Supreme Court, U. S.
FILED
SEP 13 1980
MICHAEL RODAK, JR., CLERK

No. 80-406

IN THE

Supreme Court of the United States

OCTOBER TERM, 1980

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

THE UNITED STATES OF AMERICA,
Respondent.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

ENSLEY BRANCH OF THE NATIONAL ASSOCIATION FOR THE ADVANCEMENT OF COLORED PEOPLE, *et al.*,
Respondents.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

JOHN W. MARTIN, *et al.*,
Respondent.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

LUCY WALKER, *et al.*,
Respondents.

PETITION FOR A WRIT OF CERTIORARI
to the United States Court of Appeals for the Fifth Circuit

HUBERT A. GRISSOM, JR.
DAVID P. WHITESIDE, JR.
Attorneys for Petitioners

Of Counsel:
JOHNSTON, BARTON, PROCTOR, SWEDLAW & NAFF
Twelfth Floor, Bank for Savings Building
Birmingham, Alabama 35203

St. Louis Law Printing Co., Inc., 411 No. Tenth Street 63101 314-231-4477

QUESTIONS PRESENTED

1. Whether the principle set forth in *Washington* v. *Davis,* 426 U.S. 229 (1976), that performance in a training course could be used as a criterion to show a test was job-related, set standards of test validation applicable to Title VII.

2. Whether a test having adverse racial impact which has been validated against job-relevant training may be used by an employer without a showing that success in training correlates significantly with success on the job.

3. Whether tests having adverse impact may be used to select a group of applicants having more than the minimum skills necessary to complete a job-relevant training program.

4. Whether it is appropriate, in a Title VII challenge of a test's job-relatedness, to balance the degree of validity of the test with the amount of adverse impact.

INDEX

		Page
Opinions Below		2
Jurisdiction		2
Statement of the Case		2
Reasons for Granting the Writ		6
Conclusion		20
Appendix A:	Court of Appeals Denial of Respondents' Motion for Rehearing dated June 16, 1980	A-1
Appendix B:	Court of Appeals Opinion of May 8, 1980 (review of which is sought)	A-3
Appendix C:	District Court Memorandum of January 10, 1977 (review of which is sought)	A-28
Appendix D:	District Court Order of January 10, 1977 (review of which is sought)	A-64
Appendix E:	42 U.S.C. § 2000 *et seq.*	A-66
Appendix F:	Uniform Guidelines on Employee Selection Procedures	A-76
Appendix G:	Department of Justice Guidelines on Employee Selection Procedures	A-138
Appendix H:	Equal Employment Opportunity Commission Guidelines on Employee Selection Procedures	A-171

PRECEDING PAGE WAS BLANK

CITATIONS

Cases:

Albemarle Paper Company v. Moody, 422 U.S. 405 (1975) 7,17

American Ship Building Co. v. NLRB, 380 U.S. 300 (1965) 18

Blake v. Los Angeles, 595 F.2d 1367 (9th Cir. 1979) 9

Davis v. Washington, 512 F.2d 956 (D.C. Cir. 1975) ... 9 10,11,13,15

Davis v. Washington, 348 F.Supp. 15 (D.D.C. 1972) ... 9

Espinoza v. Farah Mfg. Co., 414 U.S. 86 (1973) 17

Griggs v. Duke Power Company, 401 U.S. 424 (1971) ..7,15,18

Washington v. Davis, 426 U.S. 229 (1976) 6,9,10,11,14

Statutes and Related Authorities:

28 U.S.C. § 1254(1) 2

31 U.S.C. § 1242 2

42 U.S.C. § 1981 2

42 U.S.C. § 1983 2

42 U.S.C. § 2000(e) *et seq.* ..... passim

42 U.S.C. § 3766(c) 2

H. R. No. 914, 88th Cong., 2nd Sess. 14 (1964) 17

110 CONG. REC. 2575 (1964) 17

110 CONG. REC. 7213 (1964) 15

Other Authorities:

Department of Justice Guidelines on Employee Selection Procedures, 28 C.F.R. § 50.14 (1976) 16

E.E.O.C. Guidelines on Employee Selection Procedures, 29 C.F.R. Part 1607 (1966) 16

D. FUTRANSKY, RELATION OF D.C. POLICE ENTRANCE TESTS SCORES TO RECRUIT SCHOOL PERFORMANCE AND JOB PERFORMANCES OF WHITE AND NEGRO POLICEMEN (1967) 14

SCHMIDT, HUNTER AND URRY, STATISTICAL POWER IN CRITERION-RELATED VALIDATION STUDIES (1975) ... 16

Uniform Guidelines on Employee Selection Procedures, 5 C.F.R. § 300.103(c) (1978) 16

No.

IN THE

Supreme Court of the United States

OCTOBER TERM, 1980

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

THE UNITED STATES OF AMERICA,
Respondent.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

ENSLEY BRANCH OF THE NATIONAL ASSOCIATION
FOR THE ADVANCEMENT OF COLORED PEOPLE, *et al.*,
Respondents.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

JOHN W. MARTIN, *et al.*,
Respondent.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

versus

LUCY WALKER, *et al.*,
Respondents.

PETITION FOR A WRIT OF CERTIORARI
to the United States Court of Appeals
for the Fifth Circuit

Petitioner prays that a writ of certiorari issue to review the judgment of the United States Court of Appeals for the Fifth Circuit entered in this case on May 8, 1980.

OPINIONS BELOW

The opinion of the United States Court of Appeals for the Fifth Circuit is reported at 616 F.2d 812. That opinion affirmed in part and reversed in part the order and opinion of the United States District Court for the Northern District of Alabama dated January 10, 1977. The District Court's opinion and order are reported at 13 EPD ¶ 11,504. These opinions and orders are reproduced in Appendices A through D.

JURISDICTION

The judgment of the court of appeals was rendered on May 8, 1980. The respondents' timely petition for rehearing was denied on June 16, 1980. The jurisdiction of this Court is invoked under 28 U.S.C. § 1254(1).

Jurisdiction below was based on alleged violations of the Fourteenth Amendment, 42 U.S.C. §§ 1981, 1983, 2000(e) *et seq.* (Title VII), the Omnibus Crime Control and Safe Streets Act of 1968, as amended, 42 U.S.C. § 3766(c), and the State and Local Fiscal Assistance Act of 1972, as amended, 31 U.S.C. § 1242.

STATEMENT OF THE CASE

The Personnel Board of Jefferson County, Alabama, is the civil service employment agency for Jefferson County, the cities and towns in Jefferson County (including Birmingham, Alabama), and most governmental boards and agencies in the county. The Personnel Board is responsible for recruiting, screening and certifying qualified applicants for jobs within the civil service system. As part of its certification program, the Board develops its own examinations or purchases professionally developed examinations for use in screening and selecting qualified candidates for civil service jobs.

In 1965 the Personnel Board procured professionally developed police officer and firefighter examinations from the International Personnel Management Association (IPMA).

After the decision was made to use the IPMA examinations, the Board initiated in-house validation studies to insure that these entrance exams were job-related. Prior to the time Title VII was made applicable to state and local governments, the Personnel Board contracted with independent industrial psychologists to perform separate validity studies of the police officer and firefighter entrance examinations. The independent experts subsequently concluded that the police officer exam correlated significantly with performance in the police academy and performance on the job. The experts further concluded that the firefighter exam correlated significantly with performance in the firefighter academy and performance on the job during the first three years, the "apprentice" years.

On January 4, 1974, the Ensley Branch of the N.A.A.C.P. filed a suit against the Personnel Board and other government entities and individuals in Jefferson County complaining that the defendants were engaged in discriminatory hiring practices. Three days later John W. Martin and other named plaintiffs filed a similar suit against the Personnel Board, the City of Birmingham, and Jefferson County. In May, 1975, the United States of America brought suit against the Personnel Board and other governmental entities within Jefferson County alleging a pattern or practice of discrimination against blacks and women. Finally, on February 20, 1976, Lucy Walker filed suit against the Board and the Jefferson County Nursing Home. All four cases were consolidated for trial.

In December, 1976, trial was held on the limited issue of whether the Board's use of entrance examinations for the jobs of police officer and firefighter violated Title VII of the Civil Rights Act of 1964, as amended, (42 U.S.C. § 2000(e), *et seq.*), or other laws. The trial court reserved all other issues raised by the four complaints for a future hearing.

At trial, the Board's independent experts testified that performance on both the firefighter and police officer examinations

correlated significantly with performance in the training academies. The experts further testified that performance on the police officer examination correlated significantly with performance of police officers on the job. Appropriate police officers and firefighters testified that the training in both academies was critical and essential to the proper performance of firefighters and police officers. They further testified that these knowledges and skills could not be satisfactorily learned except through academy training.

Some of the district court's more important findings were:

(1) State law requires the Personnel Board to conduct examinations to "fairly test the relative capacity and fitness" of applicants for positions with local governmental agencies;[1]

(2) The Personnel Board used two objective, professionally developed tests to certify a pool of applicants as eligible for selection by various local hiring agencies for jobs as police officers and firefighters;[2]

(3) No other tests of equal or better validity were available for use by the Personnel Board;[3]

(4) The Personnel Board did not intentionally discriminate against blacks;[4]

---

[1] 13 EPD 6794, App. C, page A-28.

[2] *Id.* at 6795, App. C, page A-29.

[3] *Id.* at 6806-07, App. C, page A-55.

[4] The court said: "It is *clear* that the Personnel Board . . . has not intentionally discriminated against blacks." *Id* at 6795, App. C, page A-30 (emphasis added). The plaintiffs dropped all challenges of sex discrimination on these exams the day trial began. *Id.* at 6808-07 n.5, App. C, page A-58.

(5) The Personnel Board had conducted a bona fide affirmative action program[5] and had promptly adopted changes in its certification procedures recommended by its qualified independent consultants;[6]

(6) The Personnel Board's studies for validating its use of the tests were found to satisfy appropriate methodological and documentation requirements;[7]

(7) The validation studies demonstrated positive, statistically significant correlations between scores achieved on the tests and the level of success achieved by employees in subsequent training programs, which the court found were intended to and did "provide skills or knowledge needed for performance of the job";[8]

(8) The validation studies demonstrated a positive, statistically significant correlation between scores on the police officer

---

[5] Since 1965, the Board "not only sought to provide nondiscriminatory opportunities for black applicants, but also attempted, within the limits of its statutory duties, to rectify the racial imbalances in local government employment." 13 EPD 6795, App. C, page A-30. "Since 1965 the Board, with the cooperation of local civic groups and some of the employing agencies, has been actively engaged in recruitment efforts to attract black applicants." *Id.* at 6795, App. C, page A-31.

[6] The Board conducted various studies as to the validity, fairness and usefulness of the tests it administered. In 1974, the Board "immediately" adjusted its testing program, when advised by its experts that the use of a modified scoring key for the policeman test would increase the validity of that test for black applicants. *Id.* at 6795, App. C, page A-31.

[7] *Id.* at 6798, App. C, page A-35.

[8] *Id.* at 6799, 6802, App. C, pages A-38—A-39, A-45. The court referred to the "significant relationship" of the test results to grades in the training academies, and noted that "this relationship cannot be doubted." *Id.* at 6802, App. C, page A-45.

test and an overall job performance criterion the court found to be valid;[9] and

(9) There was no showing that the test had any differential validity for minority applicants.[10]

Nevertheless, the district court held that performance in the respective academies, as shown by academy grades, could not be used as a criterion measure to correlate with performance on the entrance examinations; and that the tests did not show "operational utility" and could not be used to rank blacks and whites as a group in light of the small correlations and great adverse impact on blacks.[11]

The court of appeals affirmed, holding where tests were used to select candidates with more than a minimum amount of knowledge necessary to complete job-related training, absent a further showing of a significant correlation between the degree of performance in the academy and on the job, the use of the test would violate Title VII.

REASONS FOR GRANTING THE WRIT

This case raises questions of great importance about the permissive scope of the use of testing under Title VII. Both the district court and the court of appeals ignored the standard set forth by this Court in *Washington* v. *Davis* that an entrance exam will be deemed to be job-related if it predicts success in a

---

[9] 13 EPD 6799, 6802-03, App. C, pages A-40, A-45—A-48.

[10] The court's detailed analysis of fairness demonstrated that the performance of blacks was, if anything, over-predicted by the policeman test. The court concluded that this test's use was "not to be criticized on the basis of differential validity inquiries." *Id.* at 6802, App. C, page A-44. The court found that it was not feasible to conduct such inquiries with respect to the firefighter test. *Id.*

[11] *Id.* at 6806-07, App. C, page A-54.

job-relevant training program. Further, the court of appeals held that absent an additional validation study showing a correlation between successful completion of the training academy curriculum and performance on the job, an employer may not use a test to select persons other than those minimally qualified to complete academy training. Finally, the lower courts ruled that a test with a low validity coefficient, but which had a high degree of adverse impact, could not be used even though it was shown to be job-related. The Board submits that these rulings are incorrect. Moreover, if permitted to stand, the decisions of the lower courts can be expected to chill the use of objective tests to select the best qualified persons for jobs not only in the Fifth Circuit but throughout the country.

1. The Lower Courts, in Holding that Success in the Training Academy, as Measured by Relative Standing, Was Not a Valid Criterion Measure, Ignored the Case of *Washington* v. *Davis*, 426 U.S. 229 (1976).

Most of the testing cases before this Court have involved "quick and dirty" validation efforts which attempted to justify poor or ill-conceived tests or made no effort to show the tests were job-related. See *Albemarle Paper Company* v. *Moody*, 422 U.S. 405 (1975); *Griggs* v. *Duke Power Company*, 401 U.S. 424 (1971). This is not one of those cases. In 1965, almost seven years before Title VII was made applicable to state and local governmental agencies, the Personnel Board began its efforts to find a job-related selection device for police officers and firefighters. After examining numerous professionally prepared tests, the Board selected the professionally developed IPMA 10-C test to use as their entrance exam for police officer and the professionally developed IPMA 20-B test for use as the entrance exam for firefighter. Validation efforts began in early 1966. In 1970, four years before the first suit was filed and two years before Title VII became applicable to the Personnel Board, initial studies indicated that the police officer and firefighter tests

used by the Board had validity. Outside consultants were hired three years before the Department of Justice brought suit.

Both the 10-C and 20-B selection devices predict success in the training academy (as measured by overall average grade in the academy) beyond *.01* level of confidence.[12] Uncontradicted testimony showed that the fire and police academies taught knowledges and skills that were essential to adequate performance of a police officer or firefighter and that the skills could not be learned on the job. The academy directors and the outside consultants testified that recruits who excelled in the academies, as a group, made better police officers and firefighters. The district court noted that even though the tests were correlated with overall academy grades, because there were no job incentives (benefits, etc.) to perform well in the academies, the use of those academy grades as criterion measures for correlation with test scores was inappropriate.[13]

Furthermore, the academy grades were based in part on paper and pencil tests, a "fault" that the district court used to support its conclusion that academy grades could not be used as a criterion measure.[14] The court then stated:

> [I]t is the court's conclusion that *relative* standing in the academies, as distinguished from successful completion of academy training, is not an appropriate criterion unless it also be demonstrated that those academy grades are themselves valid predictors of job performance.[15]

---

[12] A level of confidence of .05 is sufficient to show significance under all testing guidelines. Significance at $p < .01$ exceeds these minimum requirements.

[13] 13 EPD 6799, App. C, page A-38.

[14] *Id.*

[15] *Id.* at 6802, App. C, page A-44.

The court of appeals affirmed, stating that the district court's findings were not clearly erroneous.[16]

The lower court decisions that the correlations between performance on the entrance exams and performance in the academies were suspect because of the use of academy grades[17] disregard this Court's decision in *Washington* v. *Davis,* 426 U.S. 229 (1976).

Washington v. *Davis* involved a challenge of the hiring practices of the Metropolitan Police Department of Washington, D.C. Pursuant to a motion for summary judgment, the district court held that performance on the entrance exam was "reasonably and directly related to the" academy training performance and use of the test was permissible despite the large failure rate of blacks as a group. *Davis* v. *Washington,* 348 F.Supp. 15, 17 (D.D.C. 1972). The Court of Appeals for the District of Columbia reversed the trial court. The D.C. Circuit held that the use of academy scores as the sole criterion measurement was impermissible for validation of a test which had a disproportionate impact on blacks. *Davis* v. *Washington,* 512 F.2d 956 (D.C. Cir. 1975). This Court reversed the court of appeals and held that use of the test did not violate the Constitution or any statute, including Title VII.[18] *Washington* v. *Davis,* 426 U.S. 229 (1976).

---

[16] 616 F.2d at 818, App. B, page A-13. The Board submits that the clearly erroneous doctrine applies to findings of facts and not conclusions of law.

[17] Academy grades, as used by the independent consultants for their validation studies, consisted of a combination of written and performance test scores.

[18] While *Washington* v. *Davis* was not a Title VII case, the testing standards applied in *Washington* v. *Davis* were "similar to those . . . under Title VII." 426 U.S. at 249. There has been some question about the applicability of *Washington* v. *Davis* to Title VII cases. See, e.g., *Blake* v. *Los Angeles,* 595 F.2d 1367, 1378 (9th Cir. 1979). The Board respectfully suggests that the application of *Davis* standards to Title VII testing cases may be an additional issue the Court may want to address if it grants the Board's petition for a writ of certiorari.

The facts of *Washington* v. *Davis,* are amazingly similar to the facts of this case. The entrance test used to select Washington police officers was Test 21, which was used to test verbal ability.[19] *Davis* v. *Washington,* 512 F.2d at 958, 348 F.Supp. at 16. A passing point of at least 40 correct answers out of 80 was set for Test 21.[20] *Id.* at 966. Approximately fifty-seven percent of the blacks taking Test 21 failed, while only thirteen percent of the whites failed.[21] *Id.* at 958.

The similarities between *Washington* and the Jefferson County Personnel Board do not end with the entrance exam. Like the case at bar, the Washington police academy never had anyone fail the academy, and since a minimum score of 70 was required to pass the academy, the range of grades was restricted, with almost three-quarters of all recruits having grades between 85 and 95. *Davis* v. *Washington,* 512 F.2d at 962, n. 38; 348 F.Supp. at 16. A validation study indicated that performance on the entrance exam used to select Washington police officers correlated significantly with *written* examinations given in the training academy.[22] The *Davis* court of appeals, looking at the validation study, stated:

---

[19] The 10-C and 20-B exams were designed specifically for police officers and firefighters, and measured verbal ability. 616 F.2d at 816 n.9, App. B, page A-8.

[20] The lowest passing point on the 10-C was 31 correct answers out of 80.

[21] Fifty-two percent of the blacks failed the 10-C examination, while ten percent of the whites failed. 13 EPD 6796 App. C, page A-33.

[22] The *Davis* validation study never made any attempt to correlate performance on the entrance exam with successful *completion* of the academy.

> We think this evidence tends to prove nothing more than that a written aptitude test will accurately predict performance on a second round of written examinations . . . 512 F.2d at 962.

The court of appeals continued:

> All entrants into Recruit School pass the final examinations with a grade of 70 or above; if a particular candidate has difficulty, he is given assistance until he succeeds in passing the examinations. The validity study revealed that persons with high Test 21 scores are more likely to achieve a final average exceeding 85 in Recruit School, but there is no evidence to support the proposition that a candidate with an average below 85 is more difficult to train or will not be as good a police officer as a candidate with an average over 85. Moreover, since applicants who scored below 40 on Test 21 have never been admitted to Recruit School, the validity study expressed no conclusion regarding the likely performance in Recruit School of Test 21 failures. For these reasons, and because of the departmental policy that nobody fail Recruit School, appellees have not shown that that admission of applicants who score below 40 on Test 21 into the Recruit School would necessitate expanded training time or produce Recruit School failures. We might add that the Recruit School averages apparently have not been used by the Department for any purpose other than the attempt to validate Test 21 in this case. 512 F.2d at 963. (footnotes omitted)

For those reasons the D.C. Circuit held that use of academy grades as a criterion measure was improper, and remanded the case to the district court to determine whether the use of Test 21 was warranted. This Court reversed the court of appeals and agreed with the district court that a positive relationship between performance on an entrance examination and in a training academy was sufficient to validate an entrance examination. *Washington* v. *Davis,* 426 U.S. 229, 248-53 (1976).

The lower courts, in the case at bar, committed the same errors. They ignored the uncontraverted testimony of seven witnesses, who were experts in their fields and who were personally acquainted with the practices of the police and fire academies, that the academies taught skills and knowledges that were not just job-relevant, but were essential to adequate performance of a police officer or firefighter. Testimony that the skills and knowledges taught in the academy could not be learned on the job was also ignored. Instead, the lower courts reasoned because correlations between the entrance examinations and academy scores were based, in part, on performance on *written* examinations,[23] and because there were no incentives for superior performance in the academies,[24] the use of academy grades for a criterion-measure was impermissible. The lower courts in this case and the court of appeals in *Washington* v. *Davis* used identical reasoning to find that academy performance could not be used as a criterion measure. As this Court said, such reasoning is error and is due to be reversed.

---

[23] The trial court overlooked the fact that written examinations test the knowledge of theories which are essential to satisfactory performance as a firefighter or police officer. For example, a firefighter must know not only that a fire is burning, but he must know what type of fire it is and what chemicals must be used to extinguish it.

An additional benefit of a written objective test is that it is less subject to personal bias and prejudices. Such bias and prejudice might accompany other testing procedures. There is less of a chance for discrimination against minorities with the use of objective written test than there would be with the use of subjective selection devices.

[24] While there were no "rewards" or "punishments" based on performance in the academy, this does not eliminate the employer's need and desire to have all trainees perform well in the academy. Furthermore, the academy directors testified that those that performed well in the academies, as a group, made better police officers and firefighters.

This Court should take the opportunity to resolve, once and for all, the issue of whether a test which predicts success in job-relevant training meets the job-related standards of Title VII. We respectfully submit that the district court's imposition of additional requirements beyond the showing that the test predicts success in the academy is unwarranted and certainly not contemplated by the framers of Title VII. The imposition of these additional standards discourages employers from developing their own training programs. This may result in the employer turning to additional certification or educational requirements which conceivably would have more impact upon protected groups than the use of written tests.

2. This Court Should Permit the Use of a Test, Which Is Shown to Be Job-Related, to Select Those Persons Who Are Best Qualified for the Job.

The Fifth Circuit read *Davis* to require an additional step in the validation process. This additional step was that the Board's tests, after a showing of correlations between performance on the tests and performance in job-relevant training, could be used only to predict the *minimum* skills to *complete* the training program.[23] The court went on to state:

> *Washington* v. *Davis* holds that a selection device may be validated if it is shown to predict whether an applicant has a *minimum* amount of reading and verbal skills necessary to *complete* a job-relevant training program. 616 F.2d 821, App. B, page A-20 (Emphasis added).

The Board submits that the Fifth Circuit not only misconstrued the facts of *Davis,* but also established a new standard of testing which could have severe impact on the legitimate goal of employing the best qualified person for the job.

---

[23] Id. at 818, App. B, page A-20.

The facts of *Davis* do not show that the *Davis* test was selecting those police officer candidates who had the *minimum* skills to complete job-relevant training. The record shows that the validation study used by the Metropolitan Police Department indicated that the number of correct answers necessary to pass Test 21 could be lowered from 40 to 35 without producing academy failures.[26] Yet the passing point was never lowered. Thus, the test approved in *Davis* did not select those who had minimum skills. Likewise, the record shows there was no prediction that those who failed Test 21 could not complete the academy training. As the court of appeals in *Davis* stated:

> The validity study revealed that persons with high Test 21 scores are more likely to achieve a final average exceeding 85 in the Recruit School, but there is no evidence to support the proposition that a candidate with an average below 85 is more difficult to train or will not be as good a police officer as the candidate over 85. Moreover, since applicants who scored below 40 on Test 21 have never been admitted to Recruit School, the validity study expressed no conclusion regarding the likely performance in Recruit School of Test 21 failures. For these reasons, and because of the departmental policy that nobody fail Recruit School, appellees have not shown that the admission of applicants who score below 40 on Test 21 into the Recruit School would necessitate expanded training time or produce Recruit School failures. 512 F.2d at 963. (footnotes omitted)

---

[26] D. FUTRANSKY, RELATION OF D.C. POLICE ENTRANCE TESTS SCORES TO RECRUIT SCHOOL PERFORMANCE AND JOB PERFORMANCES OF WHITE AND NEGRO POLICEMEN 2 (Report published by United States Civil Service Commission) (1967), (hereinafter Futransky Study). The Futransky Study was reprinted in the Appendix in *Washington* v. *Davis,* at page 99, *et seq.* and in the Justice Department Supplemental Appendix in the appeal to the Fifth Circuit of the case at bar, at page 2, *et seq.*

The Fifth Circuit's interpretation that *Davis* only permits the use of tests which predict success in training to select those candidates with the *minimum* qualifications to *complete* training is unsupported by the facts in *Davis*.

The Fifth Circuit's opinion does much more than misconstrue the facts of *Davis*. It establishes a new requirement in the testing area that goes beyond the job-related standards of Title VII. As we read the Fifth Circuit's opinion, it permits the selection of only those people with the *minimum* qualifications for the job. It prohibits the selection of employees, who, through performance on a job-related test, have demonstrated that they will learn the necessary skills and knowledge of the job more readily. This Court has made it clear that an employer may select the *best qualified* employee without violating Title VII. *Griggs* v. *Duke Power Co.*, 401 U.S. 424, 430-31 (1971). See, also, 110 CONG. REC. 7213 (1964).

Title VII permits an employer to select those employees who are best qualified for the job. Title VII does not require an employer to hire anyone who is marginally competent but instead encourages employers to select those people who are best qualified for the job without considering their race, color, sex, creed or national origin. The Fifth Circuit rejected this theory of selecting the best qualified person for the job and would require employers to use tests only as screening devices to eliminate those people who are clearly and demonstratively unfit to learn the duties of a given job.

3. Comparing the Magnitude of Significance with Degree of Impact Should Have No Bearing on Use of the Test.

The Department of Justice Guidelines[27] provides:

> Reliance upon a selection procedure which is significantly related to a criterion measure, but which is based upon a study involving a large number of subjects[28] and has a low correlation coefficient will be subject to close review if it has a large adverse impact. § 12(b)(5).

---

[27] *Department of Justice Guidelines on Employee Selection Procedures,* 28 C.F.R. § 50.14 (1976). The D.O.J. Guidelines are attached as Appendix G. The D.O.J. Guidelines were promulgated just before trial and cited by the district court in its opinion. During the validation work by the Board and its outside consultants, the *E.E.O.C. Guidelines on Employee Selection Procedures,* 29 C.F.R. Part 1607, were the only testing guidelines published by a government agency. The E.E.O.C. Guidelines are attached as Appendix H. After the district court rendered its opinion the E.E.O.C. and the Justice Department together with other enforcement agencies promulgated the *Uniform Guidelines on Employee Selection Procedures,* attached as Appendix F.

The E.E.O.C. Guidelines differ markedly from the D.O.J. and Uniform Guidelines. The Uniform Guidelines represent an evolutionary step in the D.O.J. Guidelines.

The same theory of comparing the magnitude of impact with the "amount" of validity was carried from the D.O.J. Guidelines to the Uniform Guidelines. *Compare,* § 12(b)(5) App. G, with § 14 B(6) App. H.

[28] The Board's studies were based on samples of less than 200 people. According to generally accepted psychological testing authorities, this is only a moderate size sample. Two hundred subjects is not a large number for criterion-related validation studies. SCHMIDT, HUNTER AND URRY, STATISTICAL POWER IN CRITERION-RELATED VALIDATION STUDIES 6-9, and 12-13 (United States Civil Service Commission, Professional Series 75-7) (1975).

The district court seemed to be reaching for the same proposition in its practical utility analysis. The court conceded that the 10-C police officer exam predicted success on the job. Once job-relevance was demonstrated, the court should have ended its analysis. The district court however then engaged in a lengthy analysis of the "practical utility" of the exam. Apparently this practical utility analysis was an attempt to balance the impact of the exam against its validity. This was certainly the position advanced by Dr. Ramsey, the plaintiff's expert, when he conceded that the police officer exam predicted success in the academy *and on the job,* but *because of the impact of the exam on blacks,* he could not support its use. The Board submits that the attempts by the Guidelines, the district court and the plaintiff's expert to balance validity with impact are inappropriate and unsound testing theories and are akin to mixing apples and oranges.

Although the testing guidelines are entitled to "great deference", the guidelines are "not administrative regulations promulgated pursuant to formal procedures established by the Congress." *Albermarle Paper Co.* v. *Moody,* 422 U.S. 405, 431 (1975). Congress has withheld from the EEOC, for example, the authority to promulgate substantive rules in this area. 42 U.S.C. 2000(e)-12(a).[29] The guidelines have no legislative force, and are at most, interpretive rules. Interpretive rules are, of course, not binding upon the courts. *Espinoza* v. *Farah Mfg. Co.,* 414 U.S. 86 (1973).

This is particularly true of the Department of Justice Guidelines. The Justice Department is not a regulatory agency

---

[29]Such power would have been granted under § 714(a) of the proposed bill as reported by the House Judiciary Committee, see H. R. No. 914, 88th Cong., 2nd Sess. 14, 31 (1964), but was limited to *procedural* rule making by a floor amendment sponsored by Representative Celler. 110 CONG. REC. 2575 (Feb. 8, 1964).

engaged in the quasi-legislative and quasi-judicial administration of complex regulation pursuant to delegated authority. It is not charged with the execution of a statute; it is the courts that are so charged. The Justice Department is an advocate with both the power and the duty to come into the courts to achieve its goals like any other advocate and as such its interpretations are particularly open to rebuttal. The deference owed to even an "expert tribunal" —much less that owed to an advocacy agency — "cannot be allowed to slip into a judicial inertia which results in the unauthorized assumption by an agency of major policy decisions made by Congress." *American Ship Building Co.* v. *NLRB,* 380 U.S. 300, 318 (1965).

Title VII allows the hiring of the most qualified applicant without regard to race, color, creed, sex, or national origin. 42 U.S.C. 2000e-2(a), (j). Congress has not commanded that less qualified individuals be preferred over those with better qualifications simply because they are of a minority group. *Griggs* v. *Duke Power Co.,* 401 U.S. 424 (1971). If entrance examination, results have significant correlations with job (as the 10-C exam does here), that performance exam is selecting people who, as a group, are better qualified to perform the job. To add an additional requirement that the test significance be compared with impact on the job does not increase or decrease the validity of the test. The test remains valid regardless of the amount of impact.

What these balancing requirements do, however, is work against an employer's affirmative action programs. As an example, County X has a work force that is twenty percent black and over the past few years twenty percent of the applicants on a law enforcement entrance examination have been black. Fifty percent of the blacks and fifty percent of the whites taking the examination fail, twenty percent of the blacks are hired as police officers, and the test has been validated with performance on the job with a correlation coefficient of .20.

County Y has exactly the same racial composition, uses the same test, has obtained the same correlation coefficient in its validity study and had the same pass-fail rate for blacks and whites until two years ago. At that time, County Y instituted an affirmative action program whereby black applications increased to make up fifty percent of the applicant total. At the same time because of the decrease in the overall qualifications of the black group caused by the increased number of unqualified applicants who responded to the affirmative action program, the black failure rate jumped to seventy percent. County Y would have a very large impact and a very slight validity. Under the lower courts' rationale, County Y would have violated Title VII. Yet in County X the lack of adverse impact would not affect the small but significant correlation coefficient, even though the same test was used, and County X would not be in violation of Title VII. The sole cause of the difference is County Y's adoption of an affirmative action program.

The use of this example is intended to illustrate that the comparison of significance in a validation study with impact can cause great problems for those employers who institute affirmative action programs. The net result is to discourage affirmative action plans if one has a validated test. Both the Department of Justice and Uniform Guidelines, in an attempt to develop uniform testing procedures, have made the adoption of affirmative action programs a course of action fraught with risk.

If the comparison between levels of validation and amount of impact are deemed to be legitimate requirements of Title VII, additional problems could result. For example, a test which may be deemed valid one time it is given may not be deemed valid the next time it is given if judged on its impact. As stated earlier, impact can vary due to a number of outside forces including affirmative action plans. Furthermore, use of nationwide tests may be valid for some states or regions but may be invalid for others

if educational differences, for example, cause adverse impact within a particular race or group. A test's validity should be judged on its ability to select qualified applicants and not on the shifting variables of what percentage of various racial or ethnic groups take and pass the test.

In light of the significant efforts by the Personnel Board to institute affirmative action recruitment programs and in light of the lack of discriminatory intent, the Personnel Board submits that it is unfair and wholly inappropriate to balance impact of a test against statistically significant correlations between the entrance exam and performance on the job.

CONCLUSION

For the reasons stated above, the petition for a writ of certiorari should be granted.

Respectfully submitted,

HUBERT A. GRISSOM, JR.
DAVID P. WHITESIDE, JR.
Attorneys for Petitioner

Of Counsel:

JOHNSTON, BARTON, PROCTOR,
SWEDLAW & NAFF
1212 Bank for Savings Building
Birmingham, Alabama 35203
(205) 322-0616

CERTIFICATE OF SERVICE

I certify that three copies of the foregoing Petition for Certiorari have this date been duly served by mail upon all parties represented in this proceeding.

This the 12th day of September, 1980.

DAVID P. WHITESIDE, JR.

80-406

No.

Supreme Court, U. S.
FILED
SEP 13 1980
MICHAEL RODAK, JR., CLERK

IN THE

Supreme Court of the United States

OCTOBER TERM, 1980

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,
versus
THE UNITED STATES OF AMERICA,
Respondent.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,
versus
ENSLEY BRANCH OF THE NATIONAL ASSOCIATION FOR THE ADVANCEMENT OF COLORED PEOPLE, *et al.*,
Respondents.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,
versus
JOHN W. MARTIN, *et al.*,
Respondent.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,
versus
LUCY WALKER, *et al.*,
Respondents.

APPENDIX TO PETITION FOR A WRIT OF CERTIORARI
to the United States Court of Appeals for the Fifth Circuit

HUBERT A. GRISSOM, JR.
DAVID P. WHITESIDE, JR.
Attorneys for Petitioners

Of Counsel:
JOHNSTON, BARTON, PROCTOR,
SWEDLAW & NAFF
Twelfth Floor, Bank for Savings Building
Birmingham, Alabama 35203

St. Louis Law Printing Co., Inc., 411 No. Tenth Street 63101 314-231-4477

TABLE OF CONTENTS

		Page
Appendix A	Order of U.S. Court of Appeals Denying Petition For Rehearing	A-1
Appendix B	Opinion of U.S. Court of Appeals	A-3
Appendix C	Opinion of U.S. District Court	A-28
Appendix D	Order of U.S. District Court	A-64
Appendix E	Civil Rights Act of 1964 (42 U.S.C. §2000(e))	A-66
Appendix F	Uniform Guidelines on Employee Selection Procedures (1978); 28 C.F.R. §50.14 (1978), 29 C.F.R. §1607 (1979)	A-76
Appendix G	Department of Justice Guidelines on Employee Selection Procedures (28 C.F.R. §50.14 (1976))	A-138
Appendix H	E.E.O.C. Guidelines on Employment Selection Procedures (29 C.F.R. §1607 (1970))	A-171

APPENDIX A

IN THE UNITED STATES COURT OF APPEALS
FOR THE FIFTH CIRCUIT

No. 77-1819

Ensley Branch of the N.A.A.C.P., et al.,
Plaintiffs-Appellees,

versus

George Seibels, et al.,
Defendants,

Personnel Board of Jefferson County,
Defendant-Appellant.

John W. Martin, et al.,
Plaintiffs-Appellees
Cross Appellants,

versus

City of Birmingham, et al.,
Defendants,

Personnel Board of Jefferson County,
Defendant-Appellant
Cross Appellee.

United States of America,
Plaintiff-Appellee
Cross-Appellant,

versus

Jefferson County, et al.,
Defendants,

Personnel Board of Jefferson County,
Defendant-Appellant,
Cross Appellee.

Lucy Walker, et al.,
Plaintiffs-Appellees,

versus

Jefferson County Home, et al.,
Defendants,

Personnel Board of Jefferson County,
Defendant-Appellant.

Appeals from the United States District Court for the Northern District of Alabama

---

ON PETITION FOR REHEARING

Before GODBOLD, RONEY and ANDERSON, Circuit Judges.

PER CURIAM:

IT IS ORDERED that the petition for rehearing filed in the above entitled and numbered cause be and the same is hereby denied.

Dated June 16, 1980.

APPENDIX B

Ensley Branch of the N.A.A.C.P., et al.,
Plaintiffs-Appellees,

v.

George Seibels, et al.,
Defendants,

Personnel Board of Jefferson County,
Defendant-Appellant.

John W. Martin, et al.,
Plaintiffs-Appellees
Cross Appellants,

v.

City of Birmingham, et al.,
Defendants,

Personnel Board of Jefferson County,
Defendant-Appellant
Cross Appellee.

United States of America,
Plaintiff-Appellee
Cross-Appellant,

v.

Jefferson County, et al.,
Defendants,

Personnel Board of Jefferson County,
Defendant-Appellant,
Cross Appellee.

Lucy Walker, et al.,
Plaintiffs-Appellees,

v.

Jefferson County Home, et al.,
Defendants,
Personnel Board of Jefferson County,
Defendant-Appellant.

No. 77-1819

UNITED STATES COURT OF APPEALS
FIFTH CIRCUIT

May 8, 1980

Appeals from the United States District Court for the Northern District of Alabama.

Before GODBOLD, RONEY and ANDERSON, Circuit Judges.

R. LANIER ANDERSON, III, Circuit Judge:

This case is a consolidation of four separate actions brought by the United States and private parties against the Personnel Board of Jefferson County, Alabama, and other local government agencies. The district court held that the Personnel Board's use of two examinations in screening and certifying candidates for jobs as police officers and firefighters violates Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. §2000e et seq. The court ordered certain remedies, the remedies being tailored to its conclusion that use of the police officer test did not violate Title VII until April 25, 1975, and that use of the firefighter test did not violate Title VII until July 8, 1976. We agree with the district court that the Personnel Board's use of both tests abridges Title VII; however, we believe that the court

failed to make essential findings in determining the time at which liability commenced. Accordingly, we affirm in part and reverse and remand in part with instructions.

On January 4, 1974, the Ensley Branch of the National Association for the Advancement of Colored People, together with certain named individuals, for themselves and on behalf of others similarly situated, filed a complaint in the United States District Court for the Northern District of Alabama, against George Seibels (then Mayor of Birmingham, Alabama), the City of Birmingham, the members of the Personnel Board of Jefferson County, and the Personnel Director of that Board, alleging that the defendants engage in discriminatory hiring practices against blacks in violation of the Fourteenth Amendment, 42 U.S.C. §§ 1981, 1983, and 2000e et seq. (Title VII). A suit raising the same constitutional and statutory allegations was filed on January 7, 1974, by John W. Martin and other named plaintiffs against the City of Birmingham, Jefferson County, and the Personnel Board of Jefferson County.[1] On May 27, 1975, the United States brought suit against the Jefferson County Personnel Board and the municipal and other governmental jurisdictions within Jefferson County[2] alleging a pattern or practice of discriminatory employment practices against blacks and women in violation of Title VII, the Omnibus Crime Control and Safe Streets Act of 1968, as amended, 42 U.S.C. § 3766(c), the State and Local Fiscal Assistance Act of 1972, as amended, 31 U.S.C. § 1242, the Fourteenth Amendment and 42 U.S.C. § 1981. On

---

[1] Also named as defendants in the *Martin* suit are the Director of the Jefferson County Personnel Board and three Jefferson County Commissioners.

[2] Defendants in the government's suit other than the Personnel Board are Jefferson County, the County Department of Public Health, and the cities of Bessemer, Birmingham, Fairfield, Fultondale, Gardendale, Homewood, Huey Town, Midfield, Mountainbrook, Pleasant Grove, Tarrant, and Vestavia Hills.

February 20, 1976, Lucy Walker filed suit challenging the employment practices of the Jefferson County nursing home under Title VII and 42 U.S.C. § 1981. All four cases were consolidated for trial.

On December 20-22, 1976, trial was held on the merits of the limited issue of whether the two tests used by the Personnel Board to screen and rank applicants for positions as police officers and firefighters are discriminatory and violative of the constitutional or statutory rights of blacks.[3] All other issues under the complaints were reserved until a later date.

On January 10, 1977, pursuant to Fed.R.Civ.P. 54(b), the district court entered a final order on this limited issue.[4] The court held that the police and firefighter tests do not violate the Constitution,[5] but do violate Title VII. The court determined that the Title VII violation commenced on April 25, 1975, with respect to the police test and on July 8, 1976, with respect to the firefighter test. The Court ordered certain remedies to correct

---

[3] At trial, plaintiffs dropped their attack on the Office Worker test administered by the Personnel Board and also dropped their claim that the police and fire tests discriminate on the basis of sex.

[4] The district court's opinion and order is reported at 13 Empl.Prac.Dec. (E.P.D.) ¶11,504.

[5] The district court found that the Personnel Board's use of the tests in issue is not motivated by intentional discrimination. The parties do not dispute this finding. This appeal, therefore, is restricted to Title VII issues.

the discrimination caused by use of the tests since those dates.[6] From this judgment, the Personnel Board filed a notice of appeal, and the United States and the plaintiffs in the *Martin* action filed a joint notice of cross-appeal. The Personnel Board contends that the tests do not violate Title VII; the United States and the *Martin* plaintiffs contest the court's determination as to when Title VII liability commenced.

This case has been ably argued on appeal by counsel for both sides,[7] and we also have the benefit of a well-reasoned and comprehensive opinion by Judge Pointer of the court below.

FACTS

The Personnel Board of Jefferson County is required under Alabama law to administer examinations to applicants for positions with local government agencies. Two such examinations are at issue here: the 10—C test administered to applicants for positions with the fire department, and the 20—B test administered to applicants for positions with the police department.

---

[6] Specifically, the Court ordered that blacks be referred for openings on the police and firefighter forces at the rate at which they took the tests when most recently administered. To accomplish this, the Court ordered that the names of a sufficient number of blacks be added to the current police and firefighter eligibility lists so that the lists shall be representative of the racial composition of the test-takers, i.e., 28 and 14 percent black for police and firefighter lists, respectively; that, one-third of future certifications, *i.e.*, referrals from the lists for actual employment, are to be black until, considering all certifications since the relevant 1975 and 1976 dates, the numbers of certifications become representative of the racial composition of the test-takers. Thereafter, blacks are to be certified in accordance with their representation of the lists, *i.e.*, 28 and 14 percent of certifications for policemen and firefighters, respectively, will be black. Similarly, referrals from future lists will be a function of the rate at which blacks take the examinations on which the lists are based, until or unless defendants develop valid tests. 13 E.P.D. at 6807-08.

[7] The Educational Testing Service, Inc. filed a brief as *amicus curiae* urging reversal of the district court's holding that the tests violate Title VII.

Both tests were developed by the International Personnel Management Association.[8] Each test is a paper and pencil instrument consisting of 120 multiple choice questions.[9]

Applicants who pass[10] the 10—C or 20—B are placed on an eligibility list for positions with the police or fire department, and are ranked on the list in the order of their scores. If a single vacancy occurs, the three persons at the top of the list are certified to the appropriate department for final selection. If multiple vacancies occur, the number of persons certified is two more than the number of vacancies to be filled. Thus, merely passing the tests, and thereby being entered on the relevant eligibility list, is not nearly so important as obtaining a score sufficiently high to be placed high enough on the list to be actually certified to the police or fire department.

ADVERSE IMPACT

A *prima facie* Title VII case against an employment test may be built with statistics showing that use of the test has an adverse racial impact. This showing shifts to the employer the burden of proving that the test is job-related. *Albemarle Paper Co.* v. *Moody,* 422 U.S. 405, 425, 95 S.Ct. 2362, 2375, 45 L.Ed.2d 280 (1975); *Griggs* v. *Duke Power Co.,* 401 U.S. 424, 432, 91 S.Ct. 849, 854,28 L.Ed.2d 158 (1971); *Scott* v. *City of Anniston,* 597 F.2d 897, 901 (5th Cir. 1979).

The district court found that use of the 10—C and 20—B has an adverse impact upon blacks. The evidence shows that since

---

[8] The International Personnel Management Association was formerly the Public Personnel Association.

[9] Since April 10, 1974, only 80 of the 120 questions have been used in grading the 10—C test. We have examined the tests and note in passing that a very substantial number of the questions seem to be aimed primarily at verbal or mathematical ability.

[10] The raw score which constitutes a passing grade fluctuates, depending on the number of vacancies anticipated and other factors.

March 24, 1972, the date Title VII became applicable to public employers such as the Personnel Board, only 51 blacks (6.6%) have been hired out of a total of 768 blacks who took the police test. The comparable figure for whites is 455 whites hired (23.3%) out of the total of 1,953 whites who took the police test. Similarly, only 9 black firefighters have been hired, which represents 3.2% of the black applicants, as compared to 215 white firefighters hired, which represents 14.1% of the white applicants. With respect to the number of blacks and whites who passed the exam, the court found that the pass rates for blacks (48.6% for the police test and 24.2% for the firefighter test) are substantially less than the pass rates for whites (90.2% for the police test and 82.5% for the firefighter test). The Personnel Board does not in this appeal contest the district court's finding that the two tests have an adverse impact; the Board argues only that it has carried its burden of showing that the tests are job-related.

JOB-RELATEDNESS

[2] To prove the job-relatedness of the police and firefighter tests, the Personnel Board employed criterion-related validity studies.[11] The district court held that the studies failed to

---

[11] "Validation" is the process of determining whether a selection device is sufficiently job-related to comply with the requirements of Title VII. *See* Uniform Guidelines on Employee Selection Procedure [hereinafter referred to as Uniform Guidelines], 43 Fed.Reg. 38290, 38291 (August 25, 1978). There are three basic methods of validation: "'criterion' validity (demonstrated by identifying criteria that indicate successful job performance and then correlating test scores and the criteria so identified); 'construct' validity (demonstrated by examinations structured to measure the degree to which job applicants have identifiable characteristics that have been determined to be important in successful job performance); and 'content' validity (demonstrated by tests whose content closely approximate tasks to be performed on the job by the applicant)." *Washington* v. *Davis*, 426 U.S. 229, 247, 96 S.Ct. 2040, 2051, 48 L.Ed.2d 597 n.13 (1976); *See* Uniform Guidelines, § 5, 43 Fed.Reg. at 38298.

validate either test.[12]

The studies used three criterion measures: academy grades, *i.e.*, relative class rank among students who successfully complete academy training; efficiency ratings of incumbent employees (police and firefighters) by their supervisors; and experimental ratings of incumbent employees by their supervisors, based on 12 categories, the categories relating to personality characteristics, job knowledge and job-related abilities.

The Uniform Guidelines cited in this footnote were adopted on August 25, 1978, by the Equal Employment Opportunity Commission (EEOC), the Civil Service Commission (CSC), the Department of Labor (DOL), and the Department of Justice (DOJ), and are codified in the 1979 editions of 29 C.F.R. § 1607 (EEOC); 5 C.F.R. § 300.103 (c) (CSC); 41 C.F.R. § 60-3 (DOL); and 28 C.F.R. § 50.14 (DOJ). The Guidelines, effective September 25, 1978, supersede the guidelines on employee selection procedure previously issued by the EEOC and the DOJ (the DOJ Guidelines were issued jointly with the DOL and CSC). With the exception of its provision on interim use of unvalidated tests, discussed below, the Uniform Guidelines are consistent, in all matters pertinent to this case, with the EEOC and DOJ Guidelines. This opinion shall refer to the EEOC and DOJ Guidelines as well as to the Uniform Guidelines; citations to the former will be to the 1978 C.F.R., citations to the latter will be to the Federal Register.

[12] The Personnel Board, in 1970, initiated a preliminary, in-house, validation study of the 10—C police officer test. This study, developed by Eugene Williams, Chief Examiner of the Board was not designed to meet the validation requirements of Title VII and was not relied upon by the district court in assessing the validity of the 10—C.

The studies upon which the district court based its holding of no job-relatedness, and which are the subjects of our review, were conducted by Drs. William McLaurian and William Farrar, psychology professors at the University of Alabama in Birmingham. The McLaurian-Farrar studies examined both the 10—C police officer tests and the 20—B firefighter test. The final results of these studies, (the studies began in 1972) were reported to the Personnel Board on August 25, 1975, with respect to the 10—C and on July 8, 1976, with respect to the 20—B.

Academy Grades. The district court found that the scores on the 10—C and 20—B tests bear statistically significant[13] correlation to grades in the training academies and it further found that the training academies furnished skills or knowledge needed for performance on the job. However, the district court also found that academy grades are not valid predictors of job perfor-

---

[13] Explanation of a few statistical concepts is in order. We quote at length from the government's brief:

> Statistically, the degree of correlation between two variables (*e.g.*, entrance exam scores and subsequent school grades) is expressed as a 'correlation coefficient' on a scale running from +1.0 to −1.0. A perfect positive correlation (*e.g.*, entrance exam scores exactly predict subsequent school grades, with the higher exam scores predicting the best grades) would be expressed as +1.0, and a perfect negative correlation (*e.g.*, entrance exam scores exactly predict subsequent school grades, except in reverse, with the lower exam scores predicting the best grades) would be expressed as −1.0. Where the two variables had absolutely no relationship to each other, the correlation coefficient would be .0. The closer a correlation coefficient is to either +1.0 or −1.0, the 'higher the magnitude' of the correlation; and the closer it is to .0, the 'lower the magnitude.' Mueller, Schuessler & Castner, *Statistical Reasoning in Sociology,* 2d Ed., at p. 315.

Because a purely random drawing of a sample is liable to produce a correlation coefficient which is somewhat off an absolute .0, the concept of statistical significance becomes relevant. The concept is tied to the statistical theory of probability and is dependent upon the number of people in the sample. Generally, if a correlation coefficient is so low that, on the basis of the sample size involved, more than 1 in 20 random drawings could be expected to produce a correlation coefficient is considered not to be statistically significant, or simply to be the same as a correlation coefficient of .0. On the other hand, if the obtained coefficient could be expected to reoccur no more than once in 20 random drawings, it is considered statistically significant, the statistical indication for which is $p. < .05$. A correlation coefficient of the obtained magnitude which could not be expected to occur by chance more than once in 100 random drawings is expressed as $p. < .01$. Mueller, et al., pp. 394, *et seq.*

mance.[14] The court concluded from these findings that while completion of the academy is a valid criterion measure, academy grades are not. Finally, the court found that even though completion of the training academies is a valid criterion measure, neither test predicts successful completion of the academies; even extremely low test scores do not predict failure at the academies.

Efficiency Ratings. Drs. Farrar and McLaurian, the experts who conducted the validation studies, testified that the efficiency ratings are not trustworthy assessments of the employees' actual performance. Largely based on this testimony, the court held that the efficiency ratings are not a valid criterion measure.

Experimental Ratings. Although determining that the experimental ratings system is an appropriate criterion measure for determining job-relatedness, the court discovered fatal deficiencies in the relationship between the test scores and the ratings. With respect to the 20—B test, the court found that while there is a statistically significant positive correlation between test scores and the experimental ratings for firefighters having less than three years' experience, there is a significant negative correlation for firefighters having more than three years' experience, thus "suggesting that over time the lower scoring applicants may be better employees." Accordingly, the court held that the 20—B is not a valid predictor of the experimental ratings. With respect to the 10—C test, the court found that there is a statistically significant correlation between test scores and experimental ratings, but that the correlation is

---

[14] The court pointed out that "for the most part the correlation between academy grades and measures of job performance are not significant and, in the few instances where significant correlations are found, the findings are mixed, some being positive and others being negative. A negative correlation, of course, indicates that the higher the academy grades, the lower the performance ratings tend to be." 13 E.P.D. at 6802.

of very low magnitude and lacks practical significance.[15] Accordingly, the court held that the correlation between the 10—C and the experimental ratings does not validate the 10—C for operational use in screening or ranking applicants.

The Personnel Board raises numerous objections to the above findings by the district court respecting the validity studies. We have reviewed these finding under the clearly erroneous standard. *See Wade* v. *Mississippi Cooperative Extension Service,* 528 F.2d 508, 516 (5th Cir. 1976); *United States* v. *City of Chicago,* 549 F.2d 415, 429 (7th Cir.), *cert. denied,* 434 U.S. 875, 98 S.Ct. 225, 54 L.Ed.2d 155 (1977); *Bridgeport Guardians, Inc.* v. *Members of Bridgeport Civil Service Commission,* 482 F.2d 1333, 1337 (2d Cir. 1973), *cert. denied,* 421 U.S. 991, 95 S.Ct. 1997, 44 L.Ed.2d 481 (1975); *Washington* v. *Davis,* 426 U.S. 229, 256, 96 S.Ct. 2040, 2055, 48 L.Ed.2d 597 (1976) (Stevens, J., concurring). Because we are not left with a "definite and firm conviction" that the court committed a mistake in holding that the studies do not demonstrate job-relatedness, *Wade* v. *Mississippi Cooperative Extension Service, supra,* 528 F.2d at 516, we reject the Board's objections.[16]

---

[15] Practical or operational significance is required by the guidelines. Uniform Guidelines, § 14B(6), 43 Fed.Reg. at 38301; EEOC Guidelines, 29 C.F.R. § 1607.5(c)(2); DOJ Guidelines 28 C.F.R. Part 50.14 § 12B(5). Judge Pointer's finding that the 10—C lacks practical significance was reached through painstaking application of several different statistical approaches, *see* 13 E.P.D. at 6803-06; additionally, as permitted by the guidelines, the judge's inquiry into practical significance considered, *inter alia,* the degree of adverse impact of the 10—C. Id. at 6806.

[16] The Board's principal criticism of the findings below is the alleged failure of Judge Pointer to apply a "corrected coefficient," which would make the statistics relating to the police test more favorable to the Board. The Board has not convinced us that the district court failed to give proper consideration to the "corrected coefficient." A careful reading of Judge Pointer's opinion makes it apparent that the judge recognized that the "corrected coefficient" is based on several assumptions, and the judge, accordingly and proper-

APPLICATION OF WASHINGTON V. DAVIS

The Personnel Board's principal argument on appeal is based on the Supreme Court's opinion in *Washington* v. *Davis,* 426 U.S. 229, 96 S.Ct. 2040, 48 L.Ed.2d 597 (1976). The Board focuses on the district court's finding that both the police test and the firefighter test bear a statistically significant correlation to training academy grades. With that focus the Board argues that the *Davis* case compels a decision in its favor. The Board contends that the *Davis* case stands for the general proposition that a test can be validated by showing that it predicts grades in a job-relevant training program, without regard to the test's ability to predict job performance.[17] As discussed below, the

---

ly, used the "corrected coefficient" with due regard for its dependence on these assumptions. The crucial finding of the district court, which was reached after careful and comprehensive consideration of all the facts and circumstances, is that the validation study showed that the police test is predictive of better job performance, but that the magnitude of the positive prediction is so low that the test is worthless for all practical purposes. We cannot conclude that this finding is clearly erroneous.

The Board's only other significant attack on the district court's findings is its suggestion that several of the statistical approaches used by Judge Pointer in analyzing the operational utility of the police test was inappropriate. The Board has not convinced us in this regard. However, even if we should eliminate from our consideration all of the statistical approaches used by the district judge except the two approved by the Board, namely, the Taylor-Russell approach and the Anastasi approach, we still could not label as clearly erroneous the district court's finding that the police test is not appropriate for operational use.

[17] The Board contends that this interpretation of *Washington* v. *Davis* was adopted by the Supreme Court in its summary affirmance of the three-judge District Court decision in *United States* v. *South Carolina,* 445 F.Supp. 1094 (D.S.C.) *aff'd.* 434 U.S. 1026, 98 S.Ct. 756, 54 L.Ed.2d 775 (1978). The Board's reliance on *United States* v. *South Carolina* is misplaced for three reasons. (1) "[T]he precedential effect of a summary affirmance can extend no farther than 'the precise issues presented [in the jurisdictional statement required by Supreme Court Rule 15] and necessarily decided by those actions.'"

Davis case held that the test there was properly validated because it was shown to predict whether those tested had the *minimum* reading and verbal skills necessary to *complete* a job-

---

Illinois State Board of Elections v. *Socialist Workers' Party,* 440 U.S. 173, 182, 99 S.Ct. 983, 990, 59 L.Ed.2d 230 (1979); *Washington* v. *Confederated Bands and Tribes of the Yakima Indian Nation,* 439 U.S. 463, 477 n.20, 99 S.Ct. 740, 750 n.20, 58 L.Ed.2d 740 (1979); *Mandel* v. *Bradley,* 432 U.S. 173, 176, 97 S.Ct. 2238, 2240, 53 L.Ed.2d 199 (1977). We have examined the jurisdictional statement filed with the Supreme Court in *United States* v. *South Carolina* and determined that neither of the two issues presented specifically addresses the *Washington* v. *Davis* question of validation against training. (2) A summary affirmance affirms only the judgment and not the reasoning of the court below. *Illinois State Board of Elections* v. *Socialist Workers' Party, supra,* 440 U.S. at 182-83, 99 S.Ct. at 989-990; *Washington* v. *Confederated Bands and Tribes of the Yakima Indian Nation, supra,* 439 U.S. at 477 n.20, 99 S.Ct. at 750 n.20; *Mandel* v. *Bradley, supra,* 432 U.S. at 176, 97 S.Ct. at 2240. Therefore, the three-judge District Court's interpretation of *Washington* v. *Davis,* to the extent that interpretation differs from our own, is not binding on this Court. (3) *United States* v. *South Carolina* is factually distinguishable from the instant case. The State of South Carolina used minimum score requirements on the National Teachers Examination to certify and determine the pay levels of teachers within the State. A content validity study demonstrated that the content of the exam matched the content of teacher training programs in South Carolina, and that the minimum score requirement correlated to the minimum amount of knowledge necessary to effective teaching. 445 F.Supp. at 1112-14; *See also Id.* at 1107. After a certain date, examinees who did not achieve the minimum score, and thus lacked the minimum amount of knowledge to teach effectively, were not certified. Before that date, most examinees who did not achieve the minimum score were certified, but their compensation was less that those who scored the minimum. *Id.* at 1105-06 & n.12. Thus, the test in *South Carolina* was used solely as a measure of minimal competence to teach, much as the test in *Washington* v. *Davis* was used to determine a minimum level of required competence. In contrast, the instant tests were used not to determine minimum competence to perform as policemen or firefighters, but for ranking purposes unrelated to minimum competence.

relevant training program. We do not believe the *Davis*[18] rationale can be extended, as the Board urges, to the general proposition that any test can be validated by showing a relationship to training. More specifically, we reject the Board's suggested extension of the *Davis* holding to this case, where the tests were not used to ascertain the minimum skills necessary to complete job-relevant training[19], but rather were used to rank job applicants according to their test scores and to select only the highest test scorers for job placement.

In *Washington* v. *Davis,* two blacks (respondents in the Supreme Court but referred to here as plaintiffs), whose applications to become police officers in the District of Columbia Police Department had been rejected, claimed, *inter alia,* that a written test ("Test 21") used by the Department for recruiting purposes violated their rights under the due process clause of the Fifth Amendment, 42 U.S.C. § 1981, and D.C.Code § 1-320. Test 21, developed by the Civil Service Commission and administered throughout the federal service, was designed to test the verbal ability, vocabulary, reading and comprehension of potential recruits for the police department. In order to gain entry into the training program, a grade of at least 40 out of 80 was required on Test 21. Plaintiffs moved for summary judgment seeking a declaration that Test 21 unlawfully

---

[18] Some authorities suggest that *Washington* v. *Davis* has no application to Title VII because the action in *Davis* was brought under the due process clause of the Fifth Amendment, 42 U.S.C. § 1981, and D.C. Code § 1-320, and not Title VII. *See, e.g., Guardians Association of the New York City Police Department, Inc.* v. *Civil Service Commission of the City of New York,* 431 F.Supp. 526 (S.D.N.Y.), *vacated and remanded on other grounds,* 562 F.2d 38 (2nd Cir. 1977); *See also Davis,* 426 U.S. at 255, 96 S.Ct. at 2054 (Stevens, J., concurring). Since we conclude that *Washington* v. *Davis* is factually distinguishable, we do not reach the question of its application to Title VII.

[19] In this case, Judge Pointer found that neither test predicts successful completion of the training programs. See n.23, *infra.*

discriminated in violation of the Fifth Amendment. The defendants filed a counter-motion for summary judgment, asserting that plaintiffs were entitled to relief on neither constitutional nor statutory grounds.

The district court granted defendants' and denied plaintiffs' motions. *Davis* v. *Washington,* 348 F.Supp. 15 (D.D.C. 1972). The district court found, among other things, that a higher percentage of blacks failed Test 21 than whites, and that the test had not been validated to establish its reliability for measuring subsequent job performance. These findings were held sufficient to shift the burden of proof to defendants. *Id.* at 16. The district court found that defendants had met their burden by proving that "the Test is directly related to a determination of whether the applicant possesses *sufficient* skills requisite to the demands of the curriculum a recruit must master at the police academy." *Id.* at 17 (emphasis added).[20] Given its relationship to the training program, the court held that the lack of job performance validation did not defeat the test.

The Court of Appeals reversed on constitutional grounds. 512 F.2d 956 (D.C. Cir. 1975). Applying the Title VII principles developed in *Griggs* v. *Duke Power Co., supra,* 401 U.S. at 424, 91 S.Ct. at 849, to the Fifth Amendment claim, it held, *inter alia,* that without proof by defendants that Test 21 had been validated in regard to job performance, defendants had not rebutted plaintiffs' prima facie case of discriminatory impact. *Id.* at 961-65.

---

[20] Defendants' argument to the district court was that, "Undeniably, a police recruit must have the minimum reading skills in order to complete the curriculum at the Training Academy . . . Test 21 merely makes certain that a potential recruit has the minimum skills in reading and verbal ability." Memorandum in Support of Defendants Hampton, Spain and Andolsek's Motion for Summary Judgment, U.S. Supreme Court Records, Briefs at 97, FO—0097—75 (1975).

The Court of Appeals was in turn reversed by the Supreme Court. After holding that the D. C. Circuit had erred in applying Title VII standards to the Fifth Amendment claim of discrimination, the Supreme Court considered the statutory issues raised by the summary judgment motion. The Court approved the use of Test 21 to determine the minimum skills necessary for satisfactory progress in the training program. For the majority, Justice White wrote:

> The advisability of the police recruit training course informing the recruit about his upcoming job, acquainting him with its demands, and attempting to impart a modicum of required skills seems conceded. It is also apparent to us, as it was to the District Judge, that *some minimum verbal and communicative skill would be very useful, if not essential, to satisfactory progress in the training regimen.* Based on the evidence before him, the District Judge concluded that Test 21 was directly related to the requirements of the police training program and that a positive relationship between the test and training-course performance was sufficient to validate the former, wholly aside from its possible relationship to actual performance as a police officer. This conclusion of the District Judge that training-program validation may itself be sufficient is supported by regulations of the Civil Service Commission, by the opinion evidence placed before the District Judge, and by the current views of the Civil Service Commissioners who [are] parties to [this] case. Nor is the conclusion foreclosed by either *Griggs* or *Albemarle Paper Company* v. *Moody,* 422 U.S. 405, 95 S.Ct. 2362, 45 L.Ed.2d 280 (1975); and it seems to us the much more sensible consturction of the job-relatedness requirement.

426 U.S. at 250-251, 96 S.Ct. at 2052-2053 (emphasis added). Earlier in the opinion Justice White discussed the various ways a test could be validated:

> It is necessary, in addition, that they be 'validated' in terms of job performance in any one of several ways, perhaps by ascertaining the *minimum* skill, ability, or potential *necessary* for the position at issue.

426 U.S. at 247, 96 S.Ct. at 2051 (emphasis added). Similarly, Justice Stevens, concurring, stated:

> The test serves the neutral and legitimate purpose of requiring all applicants to meet a uniform *minimum* standard of literacy. Reading ability is manifestly relevant to the police function, there is no evidence that the required passing grade was set at an arbitrarily high level.

426 U.S. at 254, 96 S.Ct. at 2054 (emphasis added). And, later in his concurring opinion:

> As a matter of law, it is permissible for the police department to use a test for the purpose of predicting ability to master a training program even if the test does not otherwise predict ability to perform on the job. I regard this as a reasonable proposition and not inconsistent with the Court's prior holdings.

426 U.S. at 256, 96 S.Ct. at 2055. Further explaining his own opinion for the *Davis* majority, Justice White, in his dissent from the summary affirmance of *United States* v. *South Carolina,* 434 U.S. 1026, 98 S.Ct. 756, 54 L.Ed.2d 775 (1977), stated:

> *Washington* v. *Davis* . . . was thought by the District Court to have warranted validating the test in terms of the applicant's training rather than against job requirements; but *Washington* v. *Davis,* in this respect, held only that the test there involved, which sought to ascertain whether the applicant had the *minimum* communication skills necessary to understand the offerings in a police training course, could be used to measure eligibility to enter that program. The case did not hold that a training course, the comple-

> tion of which is required for employment, need not itself be validated in terms of job relatedness. Nor did it hold that a test that a job applicant must pass and that is designed to indicate his mastery of the materials or skills taught in the training course, can be validated without reference to the job. Tests supposedly measuring an applicant's qualifications for employement, if they have differential racial impact, must bear 'some manifest relationship to the employment in question,' *Griggs* v. *Duke Power Co.*, 401 U.S. 424, 432 [, 91 S.Ct. 849, 854, 28 L.Ed.2d 158] (1971) . . .

434 U.S. at 1027-28, 98 S.Ct. at 757 (emphasis added).

Thus, *Washington* v. *Davis* holds that a selection device may be validated if it is shown to predict whether an applicant has the minimum amount of reading and verbal skills necessary to complete a job-relevant training program. We decline the Personnel Board's invitation to extend the *Davis* rationale by holding that any test can be validated against training, without respect to the test's ability to predict job performance.[21] Such an extension would violate the requirement of job performance validation enunciated in *Griggs* and *Albemarle,* as well as the agency guidelines[22] elaborating upon that requirement.

---

[21] Other courts similarly refuse to read *Washington* v. *Davis* as establishing that selection devices may be validated for Title VII purposes without regard to job performance. *See, e.g., Blake* v. *City of Los Angeles,* 595 F.2d 1367, 1382 n.17 (9th Cir. 1979), *petition for cert. filed,* 48 U.S.L.W. 3125 (U.S. July 12, 1979) (No. 79-54); *Guardians Association of the New York City Police Department, Inc.* v. *Civil Service Commission of the City of New York,* 431 F.Supp. 526, 548-49 (S.D.N.Y.), *vacated and remanded on other grounds,* 562 F.2d 38 (2d Cir. 1977). *But see, United States* v. *Virginia,* 454 F.Supp. 1077, 1100-01 (E.D. Va. 1978).

[22] The agency guidelines require that a criteria-related validity study, such as the one at bar, show a positive relationship between the challenged test and job performance. Uniform Guidelines, § 5B, 43

Unlike the test upheld in *Washington* v. *Davis,* the tests used by the Personnel Board are not used to predict whether an applicant has the minimum amount of knowledge necessary to complete training[23]; rather, the tests are used to rank applicants according to their scores. Only those at the top of the eligibility list, those with the highest test scores, are certified for job placement. Even those who score a passing grade, and are deemed by the Personnel Board to possess the capacity to complete training, are not hired unless they are among the highest scorers. Use of a test for such ranking purposes,[24] rather than as a *Davis*-like device to screen out candidates without minimum skills, is justified only if there is evidence showing that those with a higher test score do better on the job than those with a lower test

---

Fed.Reg. at 38298. These guidelines, while not binding on the courts, are entitled to great deference. *Albemarle Paper Company* v. *Moody,* 422 U.S. 405, 431, 95 S.Ct. 2362, 2378, 45 L.Ed.2d 280 (1975); *Thomas* v. *E.I. duPont de Nemours & Company,* 574 F.2d 1324, 1331 n.9 (5th Cir. 1978); *Pettway* v. *American Cast Iron Pipe Company,* 494 F.2d 211, 221 (5th Cir. 1973), *cert. denied,* 439 U.S. 1115, 99 S.Ct. 1020, 59 L.Ed.2d 74 (1979); *United States* v. *Georgia Power Company,* 474 F.2d 906, 913 (5th Cir. 1973).

[23] Indeed, the district court found the tests inappropriate even for the limited purpose of determining whether an applicant has the minimum capacity to complete training. The court found that neither test is a valid predictor of ability to pass the training curriculum. 13 E.P.D. at 6802-03. It should also be noted that the Personnel Board has not contended on this appeal for such limited, *Davis*-sanctioned, use of the tests.

[24] The guidelines also distinguish tests used for screening from tests used for ranking. To validate the latter under a criterion-related study, "the user need show mathematical support for the proposition that persons who receive higher scores on the [selection] procedure are likely to perform better on the job." Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, 44 Fed.Reg. 11996, 12005 (Question & Answer 62).

score.[25] Such evidence is utterly lacking here. The Board's validation studies show that higher test scores do *not* predict better job performance.

Having accepted the district court's finding that the Personnel Board did not carry its burden of showing that the two tests are job-related, and having rejected the Board's suggestion that the *Davis* case can be extended to justify validation of these tests against training grades without regard to job performance, we affirm the district court's holding that the Personnel Board's use of the two tests violates Title VII. We turn next to the remedy issue and the subsidiary question of when the Board's Title VII liability commenced.

COMMENCEMENT OF LIABILITY

The Personnel Board became subject to the requirements of Title VII on March 24, 1972. Equal Employment Opportunity Act of 1972, Pub.L. No. 92-261, 86 Stat. 103. On that date, and for several years earlier[26], the Board was using the 10—C and 20—B to screen and rank applicants. However, the district court held that use of the police test did not begin to violate Title VII until April 25, 1975, and that use of the firefighter test did not constitute a violation until July 8, 1976. Those were the

---

[25] If academy grades are to be the criterion to which the tests are to be compared, the criterion itself must be shown to be a good measure of job performance. *See Blake, supra,* 595 F.2d at 1382; *Vulcan Society of the New York City Fire Department, Inc.* v. *Civil Service Commission,* 490 F.2d 387, 396 n.11 (2d Cir. 1973). "'The entire rationale of a criterion-related study requires that the criterion with which the test results are compared be a good measure of job performance.'" *James* v. *Stockham Valves and Fittings Co.,* 559 F.2d 310, 340 (5th Cir. 1977), *cert. denied,* 434 U.S. 1034, 98 S.Ct. 767, 54 L.Ed.2d 781 (1978), *quoting United States* v. *City of Chicago, supra,* 549 F.2d at 433.

[26] The Board started using the police and fire tests on August 18, 1967, and October 23, 1968, respectively.

dates on which the final results of the validation studies for the two tests (the studies were started in late 1972) were reported to the Board. The court explained:

> The preliminary reports from the consultants, made while more trustworthy measures of job performance were being developed [*i.e.*, the experimental ratings], contained signs of potential validity and recommended continued usage of the test pending the additional studies. Not until April 25, 1975, with respect to the 10—C, and July 8, 1976, with respect to the 20—B were the studies using these new criterion measures completed and reported to the Board. It was on these respective dates that, in the court's opinion, it should have been concluded that provisional use of the tests was no longer permissible. Prior thereto, the Board was, in the court's opinion, justified in continuing to use the tests (and the eligibility lists generated therefrom) in anticipation of favorable results from those studies.

13 E.P.D. at 6807.

It appears from the court's reference to "provisional use" that Judge Pointer was relying on § 1607.9 of the EEOC Guidelines in holding the Board free from liability pending its receipt of the final validation studies report.[27] Section 1607.9

---

[27] It appears also that in fixing the date of Title VII liability, the district court considered the good faith and lack of discriminatory intent of the Board in adopting and using the 10—C and 20—B. The court observed that the Board was compelled by state law to administer employment tests to screen and rank applicants; that the Board selected the 10—C and 20—B in the late 1960's "as the best tests then available, with the hope that black applicants would fare better than under previous tests"; that the Board conducted a preliminary, in-house, validity study of the tests even before it became subject to Title VII; and that, "at least since 1965 the Board has not intentionally discriminated against black applicants." 13 E.P.D. at 6806-6807. We do not question the court's finding that the Board acted in good faith and without discriminatory intent. However, to the extent that this finding influenced the court's determination as to

"authorize[s] provisional use of tests, pending new validation efforts, in certain very limited circumstances." *Albemarle Paper Co.* v. *Moody, supra,* 422 U.S. at 436, 95 S.Ct. at 2380. Section 1607.9 reads in full:

> § 1607.9 Continued use of tests.
>
> Under certain conditions, a person may be permitted to continue the use of a test which is not at the moment fully supported by the required evidence of validity. If, for example, determination of criterion-related validity in a specific setting is practicable and required but not yet obtained, the use of the test may continue: *Provided:* (a) The person can cite substantial evidence of validity as described in § 1607.7(a) and (b); and (b) he has in progress validation procedures which are designed to produce, within a reasonable time, the additional data required. It is expected also that the person may have to alter or suspend test cutoff scores so that score ranges broad enough to permit the identification of criterion-related validity will be obtained.

Section 1607.7 of the EEOC Guidelines provides that where it is not feasible to conduct a proper validation study,

> evidence from validity studies conducted in other organizations, such as that reported in test manuals and professional literature, may be considered acceptable when: (a) The studies pertain to jobs which are comparable (*i.e.*, have basically the same task elements), and (b) there are no major differences in contextual variables or sample composition which are likely to significantly affect validity.

---

when liability commenced, error was committed. Absence of discriminatory intent is no defense in a Title VII disparate impact case. *Griggs* v. *Duke Power Co., supra,* 401 U.S. at 432, 91 S.Ct. at 854; *Albemarle Paper Co.* v. *Moody, supra,* 422 U.S. at 422, 95 S.Ct. at 2373.

> Any person citing evidence from other validity studies as evidence of test validity for his own jobs must substantiate in detail job comparability and must demonstrate the absence of contextual or sample differences cited in paragraphs (a) and (b) of this section.

To justify use of the 10—C and 20—B under § 1607.9, the Personnel Board should have shown (1) that validation studies done elsewhere, as described in § 1607.7, provided substantial evidence of validity[28] and, (2) that at all relevant times the Board had in progress validation procedures designed to produce, within a reasonable time, the additional data required. The district court did not determine whether either of these conditions was satisfied. Because we believe that this determination "is a matter best decided in the first instance by the District Court," *Albermarle Paper Co.* v. *Moody, supra,* 422 U.S. at 436, 95 S.Ct. at 2380, we remand to the district court to ascertain whether the conditions were met.

If the district court, upon remand, finds that the use of the tests was permissible under § 1607.9 until the Board received the

---

[28] The first condition for interim use under § 1607.9 is that "(a) The person can cite substantial evidence of validity *as described in § 1607.7(a) and (b).*" (emphasis supplied). Thus, unless an employer cites to other validity studies which meet the requirements of § 1607.7, he has not satisfied § 1607.9(a). This condition lends objectivity to § 1607.9(a)'s substantial evidence requirement, and does not place an onerous burden on employers. *Compare, Friend* v. *Leidinger,* 446 F.Supp. 361, 370 (E.D.Va. 1977) (alternative holding sustained the City of Richmond's use of firefighter tests under § 1607.9 where tests had been proven job-related in a California validity study), *aff'd.* 588 F.2d 61 (4th Cir. 1978); *Buckner* v. *Goodyear Tire and Rubber Co.,* 339 F.Supp. 1108, 1115 (N.D.Ala. 1972), *aff'd.* (and district court opinion adopted), 476 F.2d 1287 (5th Cir. 1973) (provisional use of tests permitted where validation studies at another Goodyear plant were available). § 1607.9's incorporation of the requirements of § 1607.7 is especially reasonable in light of the fact that § 1607.9 permits use of a test, possibly under the cloak of immunity as discussed below, which has a discriminatory impact.

final results of the validation studies, and if the court also finds that the Board in good faith relied upon § 1607.9, then the district court would be correct in starting Title VII liability on April 25, 1975 (police test), and July 8, 1976 (firefighter test). The natural reading of EEOC Guideline § 1607.9 (as it was in effect at all times relevant to this proceeding[29]), especially when read in light of Title VII, § 713(b), is that an employer will be immune from liability during the period of permissible provisional use of an unvalidated test. Section 713(b) of Title VII provides a defense to an employer who complies with, and relies in good faith upon, EEOC Guidelines, such as § 1607.9.[30] Section 713(b) provides:

> In any action or proceeding based on any alleged unlawful employement practice, no person shall be subject to any

---

[29] After the dates fixed by the district court for the commencement of Title VII liability, two new sets of guidelines on employee selection procedure were issued: the Department of Justice Guidelines (effective November 17, 1976) and the Uniform Guidelines (effective September 25, 1978). *See supra,* n.11. Both the DOJ Guidelines and the Uniform Guidelines contain provisions on interim use which expressly deny an employer immunity in the event his pending validation study does not ultimately establish validity. *See* DOJ Guidelines § 5(h), 28 C.F.R. Part 50.14 ("If the additional studies do not produce the data required to demonstrate validity, the user is not relieved of or protected against any obligations arising under federal law."); Uniform Guidelines, *supra,* n.11, § 5J, 43 Fed.Reg. at 38298 ("If the study does not demonstrate validity, this provision of these guidelines for interim use shall not constitute a defense in any action, nor shall it relieve the user of any obligations arising under Federal law"). Had the DOJ or Uniform Guidelines been in effect during the pendency of the Personnel Board's validation studies, a different issue would be presented. However, at all times relevant to the instant proceedings, only the EEOC Guidelines, as they were in effect before the adoption of the Uniform Guidelines, were in force. The EEOC Guidelines do not contain a disclaimer of immunity similar to that contained in the DOJ and Uniform Guidelines.

[30] *Friend* v. *Leidinger,* 446 F.Supp. 361, 370 (E.D. Va. 1977), *aff'd.* 588 F.2d 61 (4th Cir. 1978) (Title VII, § 713(b) extends immunity during § 1607.9 interim use).

> liability or punishment for or on account of (1) the commission by such person of an unlawful employement practice if he pleads and proves that the act or omission complained of was in good faith, in conformity with, and in reliance on any written interpretation or opinion of the [EEOC].[31]

If, however, the district court, upon remand, finds that the conditions for § 1607.9 provisional use were not satisfied or that the Board did not in good faith rely upon that section while awaiting the final results of the validation studies, then the court must mark March 24, 1972, as the date of violation of Title VII, with respect to both tests. A new remedy must then be fashioned by the court to correct the discrimination caused by use of the tests from that date forward.

AFFIRMED IN PART, REVERSED IN PART AND REMANDED.

---

[31] By holding that the Board would not be subject to liability during the period it satisfied and relied upon § 1607.9, we are not, as plaintiffs contend, assuming that a grace period should be implied for public employers similar to the one-year grace period expressly accorded private employers (Title VII, § 716(a)), when Title VII was first enacted. *See Blake* v. *City of Los Angeles, supra,* 595 F.2d at 1376-77 (refusing to imply grace period for public employers). Rather, our holding is based on the immunity which Title VII, § 713(b) extends to an employer who complies with, and relies in good faith upon, the EEOC Guidelines.

APPENDIX C

Ensley Branch of the N.A.A.C.P., Plaintiff v. George Seibels, et al., Defendants. Civil Action No. 74-Z-12-S.

John W. Martin et al., Plaintiffs v. City of Birmingham et al., Defendants. Civil Action No. CA 74-Z-17-S.

United States of America, Plaintiff v. Jefferson County et al., Defendants. Civil Action No. CA 75-P-0666-S.

Lucy Walker et al., Plaintiffs v. Jefferson County Home et al., Defendants. Civil Action No. CA 76-M-2047-S.

United States District Court, Northern District of Alabama, Southern Division. January 10, 1977.

Memorandum of Opinion

POINTER, D.J.: Since 1945 the Personnel Board of Jefferson County has been charged under state law with the duty of periodically administering examinations to "fairly test the relative capacity and fitness" of applicants for positions with local governmental agencies.[1] 1940 Ala. Code Appx. §§645, *et seq.* (Recomp. 1958). Those who pass are ranked on an eligibility list in the order of their exam scores.[2] As vacancies occur, the three persons then at the top of the list are certified to the employing agency for final selection, the appointments being probationary in nature for the first twelve months.[3] An applicant's name may be removed from the eligibility list after having three times been certified and refused employment.

This litigation challenges the employment practices of the governmental agencies as discriminatory on the basis of race, color, and sex, and includes an attack upon the examinations administered by the Personnel Board. Presently at issue, following a trial held December 20-22, 1976, are the tests currently used to screen applicants for positions as police officers, deputy sheriffs,[4] and firefighters.[5]

An attack upon the police and firefighters exams is certainly understandable when one considers that, although the relevant

labor pool is over 25% black, yet on June 30, 1976, only 56 (or 6.5%) of the 860 police officers were black and only 9 (or 1.4%) of the 630 firefighters were black. These statistics may, however, be misleading for purposes of this lawsuit because they include the historical results of hiring practices employed long before passage of the Equal Employment Opportunity Act of 1972 or, indeed, before utilization of the tests under scrutiny at this time.

The principal focus should rather be upon the events of more recent years, with particular attention upon practices subsequent to March 24, 1972, when Title VII of the Civil Rights Act of 1964 was made applicable to the Personnel Board and the governmental agencies which it serves. Likewise, information as to the general labor pool in the area is of only marginal importance when one has, as we do, extensive data as to actual applicants for positions and there is no evidence that minority applications have been depressed by prior employment practices.

Adoption of the Current Tests

In late 1965, following an independent study as to why no blacks were then employed as police officers in the City of Birmingham, the Personnel Board decided to replace its police and firefighter exams with tests developed by the Public Personnel Association, now known as the International Personnel Management Association. IPMA tests were being widely used in other parts of the country and were considered by the Board as superior to other tests then available. The change was part of a multi-faceted program intended to increase black participation in governmental positions. (See Appendix C to X-342). Policeman Test 10—C and Firefighter Test 20—B have been in use since August 18, 1967, and October 23, 1968, respectively, as the screening examinations for these positions under the state-mandated selection procedure,[6] although at times other tests have been administered for experimental purposes or for validation studies. Since April 10, 1974, a modified scoring key (based

upon only 80 of the 120 test items) has been employed in grading the 10—C test for purposes of the eligibility list. This modification was made at the recommendation of qualified independent consultants who, after study, concluded that the scoring change would increase validity of the test for black applicants.

Intent

It is clear that the Personnel Board, in performing its functions as an employment agency for the various local governments, has not intentionally discriminated against blacks. Indeed, at least since 1965, the Board has not only sought to provide non-discriminatory opportunities for black applicants, but also attempted, within the limits of its statutory duties, to rectify the racial imbalances in local government employment. Some mention of these aims and efforts is appropriate.

It was the Board's hope that adoption of the tests now in issue would benefit black applicants, while nevertheless providing a fair "test of the relative capacity and fitness" of all applicants, as required by state law. Immediately, a study was undertaken to ascertain whether the IPMA policemen test, although a paper and pencil test, would correlate positively and significantly with a widely used non-verbal performance test of general intelligence, the Revised Beta Examination—and it did. As successful applicants were employed by the city of Birmingham, were trained at the police academy, and entered performance of their duties, information was incorporated into the Board's ongoing validation studies—which, while lacking sufficient blacks in the sample (only 6 in the 10—C sample) to permit full analysis, were considered by the Board as justifying further usage of the 10—C.[7] These studies are presented by the Board not as satisfying the requirements of the EEOC or Department of Justice guidelines on tests, but rather as indicating its efforts to see that its examinations were fair predictors of job performance even at a time when it was not subject to the provisions of Title VII. As already noted, when, in 1974, it was advised by

independent consultants that a modification of the scoring of the 10—C exam would improve the validity for black applicants, it immediately put that change into effect.

Since 1965 the Board, with the cooperation of local civic groups and some of the employing agencies, has been actively engaged in recruitment efforts to attract black applicants. In 1966 it began assuming the $10.00 medical examination costs for newly hired persons; and in 1967 it was successful in sponsoring legislation to eliminate the $1.50 examination fee previously required and to eliminate the priority previously given applicants who resided within an employing agency's jurisdiction.[8] It has experimented with a lowering of the raw score used to measure a "passing" grade on the exams where it could justify that approach on the basis of "supply" and "demand".

In short, in its selection, administration and use of the 10—C and 20—B tests, there has been no design or intent on the part of the Board to discriminate on the basis of race or color. However, the standard under Title VII of the Civil Rights Act of 1964 is not so limited[9]—rather, if the *operational effect* of test usage is to discriminate against blacks, then it is proscribed unless it is shown to be a "job related" requirement,[10] with a "manifest relation to the employment in question."[11] See U.S.C.A. § 2000e-2(h). In this inquiry the court is to follow the guidelines adopted by the EEOC and, more recently (November 17, 1976), by the Department of Justice (DOJ), absent some "cogent reason."[12] See *Watkins* v. *Scott Paper Co.*, 530 F.2d 1159 (CA5 1976). Also instructive are the 1974 A.P.A. Standards for Educational & Psychological Tests and the 1975 Principles for the Validation and Use of Personnel Selection Procedures of the A.P.A.'s Division 14.

Adverse Impact

Where the total selection process has an adverse impact upon a substantial racial group in the labor market, the individual

components of that process—such as a screening test—are also to be evaluated for adverse impact. DOJ Guidelines §4b. For purpose of this two-step analysis, data can be extracted from the evidence pertaining to administrations of the 10—C and 20—B tests which have been used for employment decisions after March 24, 1972.[13]

	10—C		20—B	
	Black	White	Black	White
Failing test	395	191	216	267
Passing test	373	1,762	69	1,263
Hired	51	455	9	215

According to the DOJ Guidelines, §4b, "A selection rate for any racial * * * group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded as evidence of adverse impact * * * Greater differences in selection rate would not necessarily be regarded as constituting adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority * * * candidates to be atypical of the normal pool of applicants from that group."

So far as the total selection process is concerned, one finds from the above data that the hiring rates for blacks (6.6% of the black applicants on 10—C and 3.2% of the black applicants on 20—B) are substantially less than eighty percent of the hiring rates for whites (23.3% and 14.1%, respectively). These greater differences in selection rates cannot be explained on the basis of inadequate numbers and, according to the court's calculations, are statistically significant: the o coefficient for the 10—C test is .193 and for the 20—B is .121, both of which are significant at $p < .001$.

Looking at the data pertinent to the test component of the selection process, one finds again that the pass rates for blacks

(48.6% for 10—C and 24.2% for 20—B) are substantially less than eighty percent of the pass rates for whites (90.2% and 82.5%, respectively). And again, according to the court's calculations, these greater differences in pass rates, which are based upon samples of adequate size, are statistically significant: the o coefficient for the 10—C test is .46 and for the 20—B is .48, both being significant at $p < .001$. Also of importance is the fact that, of the blacks who did pass the tests, 85.4% placed in the lower half of the initial eligibility lists for police officers and 89.9% placed in the lower half of the firefighter lists.[14]

Some concern can justifiably be expressed that the special recruiting efforts undertaken by the Personnel Board and other groups to attract black applicants—while commendable as an affirmative action to overcome racial imbalance in the police and firefighter forces—may at the same time have resulted in an atypical pool of blacks taking the test, producing distortion in the test performance of the black applicants. For example, the black applicants may have included many who were not seriously interested or motivated with respect to the jobs in question, thereby affecting their test performances. Absent, however, any hard data to support such an hypothesis or to indicate its magnitude, the court, impressed with the substantial differences in hire rates and pass rates for the two racial groups, must conclude that the overall selection procedures in effect since March 24, 1972, and as a component part thereof the tests used for those purposes, have had an adverse impact on blacks.

Validation Studies

According to EEOC Guidelines § 1607.3, "the use of any test which adversely affects hiring * * * of classes protected by Title VII constitutes discrimination unless (a) the test has been validated and evidences a high degree of utility as hereinafter described * * *." For the purpose of making such validation studies of its many tests, the Board in 1972 contracted with Drs. William E. Farrar and William A. McLaurin, Professors in the

Psychology Department of the University of Alabama at Birmingham. Both had experience with personnel selection procedures in public employment systems. Priority, but not exclusive attention was to be given to the police and firefighter tests, and their work on these tests began in late 1972. Their studies respecting the two tests continued even to the time of trial, with various reports being made in each of the years 1973, 1974, 1975 and 1976. That their work was not complete before trial does not suggest inattention; rather, it is indicative that their studies were intended to be thorough and were directed to numerous tests.[15]

Psychometric Analyses

The 10—C and 20—B tests are paper-and-pencil instruments, each consisting of 120 multiple choice items.[16] The initial concern of Drs. Farrar and McLaurin was directed to the reliability,[17] item difficulty,[18] and item discrimination[19] of the tests. The following findings were made:

	reliability		items satisfactorily difficult	items satisfactorily discriminating
	r_{sp}	r_{kr}		
10—C total (N = 479)	.95	.95	88	116
10—C black (N = 176)	.90	.90	77	106
10—C white (N = 303)	.95	.93	70	115
20—B total (N = 507)	.92	.91	64	115
20—B black (N = 108)	.85	.85	61	66
20—B white (N = 399)	.86	.86	53	108

Inquiry into reliability is a proper first step, because, while no test is perfectly reliable, a test which is not reliable is not valid for any purpose. The consultants found the reliability coefficients for both tests to be of sufficient magnitude to indicate satisfactory reliability. They did acknowledge that the methods selected for this purpose were essentially measures of internal consistency (and with the KR-20 formula, of content homo-

geneity), but apparently believed it either not feasible or not necessary to investigate error variance due to time sampling. The court agrees as to reliability and notes that possible lack of stability over time is, in a sense, mitigated by the fact that applicants may take an exam on more than one administration.

Analyses of item difficulty and discrimination have no direct bearing upon the validation studies before the court. However, they do reflect an investigation into possible modification or supplementation of the tests to improve their utility and reduce the extent of adverse impact, which is a recommended procedure.[20] See DOJ Guidelines, § 3c.

Documentation and Methodology

The EEOC Guidelines, at §§1607.5(b)(2,3,5) and 1607.6, require that various items of information (*e.g.* copies of tests, manuals, rating forms and instructions and representations of statistical data) be included in the report of the study or otherwise available for inspection. Following the 1974 A.P.A. Standards, a more extensive list of documentation requirements is specified in the DOJ Guidelines at §§4a and 13b, involving some twenty-four "essential" items and several other desirable items. The Farrar-McLaurin studies satisfy the EEOC requirements, which were the only ones in effect when their studies were conducted and (so far as then feasible) completed and, indeed, when supplemented by evidence presented immediately before and during trial, they also substantially satisfy the DOJ requirements, which became effective on November 23, 1976.[21]

The studies include presentations of the following statistics:

For the 10—C test:

- intercorrelation coefficients, r and r_c, for 109 Birmingham police officers (without separation by race) respecting their 10—C scores, police academy scores (school average and course grades) and latest efficiency ratings.

● means, standard deviations, and *t* tests for difference in means for 10—C scores of 38 black and 101 white Birmingham police officers.

● means, standard deviations, and r coefficients for 59 Birmingham police officers (without separation by race) respecting their 10—C scores, academy scores (average and courses) and latest efficiency ratings.

● means, standard deviations, r coefficients, and *t* tests for the following:

● ● 20 black and 76 white Birmingham police officers respecting their 10—C scores, academy averages, and latest efficiency ratings.

● ● 8 black and 140 white Birmingham police officers respecting their 10—C scores, academy scores (average and courses), latest efficiency ratings (overall and by sub-parts), and experimental ratings weighted average and by components.

● ● 49 black and 140 white police officers respecting their 10—C scores and academy scores (average and courses). (Also included are data for analyzing significance of differences in correlation coefficients through z transformations.)

● ● 83 Jefferson County deputy sheriffs (without separation by race) respecting their 10—C scores and academy averages.

● ● 77 police officers (without separation by race) from other cities served by the Personnel Board respecting their 10—C scores and academy averages.

For the 20—B test: means, standard deviations, and r coefficients for the following:

● 162 Birmingham firefighters (without separation by race) respecting the 20—B scores, training academy average, and latest efficiency ratings (overall and by sub-parts).

- 196 Birmingham firefighters (without separation by race) respecting their 20—B scores, academy averages, latest efficiency ratings, and experimental ratings (overall and by components). Statistics are reported separately for short-tenure and long-tenure firefighters, using three years of experience as the point of division.

Statistics found to be significant at $p < .05$ and $p < .01$ are so identified in the report.

Some commment should be made about selection and composition of the different samples. Each sample contained all the persons for whom, so far as was known at the time by the consultants, the data needed for that study was available. The different studies were, however, conducted over a period of several years as either the need was recognized or the particular inquiry became technically feasible; and during the time intervals the work force had changed. Some of the studies involved concern with additional factors (*e.g.*, performance on the Raven and PAS tests, which have been under consideration for use as supplemental or alternative screening instruments), for whom the data existed only for a limited number of applicants or employees. The result is that a particular sample may contain some, but not necessarily all, of the persons in another sample and may also contain some persons who were not in the other sample. This lack of autonomy or consistency complicates somewhat the process of analysis, but, under the circumstances, is acceptable. There is no hint of contrivance in selection of the samples or of lack of representatives of the sample subjects.[22]

Criteria

The several criterion-measures (academy grades, efficiency ratings, and experimental ratings) have certain common factors: (1) None appears to be "contaminated" (*i.e.*, affected by knowledge by the rater or scorer of prior score on the 10—C or 20—B). (2) None has been subjected to special statistical

scrutiny to detect or control possible bias among raters or graders.[23] (3) None has been subjected to special statistical scrutiny for reliability.[24] (4) Each has been analyzed by the consultants for relevancy (*i.e.*, the extent to which it may be considered as a measure of critical or important work behaviors). Each of the measures has, of course, its own special characteristics and limitations, which will be described separately. It must be emphasized that, in a criterion-related validation study, one is attempting to estimate the extent to which a score on a "predictor" (*e.g.*, 10—C test) can predict job performance (*i.e.*, as a police officer) through evaluating its ability to predict scores or ratings on a "criterion" (*e.g.*, academy average)—and hence the study is subject to any limitations which those same criteria have in either predicting or assessing job performance.

(1) Academy Grades.—Where, as here, new employees are required to complete special training before performing their duties, successful completion of that training may properly be used as a criterion-measure, if, that is, the training is intended to, and does, provide skills or knowledge needed for performance of the job. Based upon the evidence presented, including testimony of the directors of the Birmingham police and fire academies, the court finds that the two schools do serve that purpose and function.

Relative standing or ranking among students who successfully complete such training is not, however, as such, an appropriate criterion.[25] Rather, to be relevant as a criterion, such measures must be shown, empirically or otherwise, to be themselves appropriate predictors of job performance. This, in essence, means a two-step correlation study; and, in a situation where one has data on test scores, academy grades, and measures of job performance for the same group of persons, the more direct inquiry (correlation between test scores and measures of job performance) would be preferred to the two-step approach. Grades on particular courses in the academy must also be

analyzed for compatibility with findings respecting grades on other courses.

So far as the evidence indicates, academy grades—provided they are passing scores—have no impact on job opportunities, benefits, etc. If this be the case, then, while helpful in preventing "contamination" during validity studies, academy grades are likely to be influenced by motivational considerations not present in actual job performance. The emphasis in the academies on paper-and-pencil multiple choice items, while providing objectivity, may also reflect a relationship to the paper-and-pencil screening exam not found in job performance. These concerns should cause one to be cautious in making non-empirical judgments about the usefulness of relative academy grades as a criterion-measure.

(2) Efficiency Ratings.—The efficiency ratings given periodically on all employees by their supervisors are direct and, ostensibly, appropriate measures of job performance. Drs. Farrar and McLaurin have, however, acknowledged that these ratings are not trustworthy assessments of the employees' actual performance. In addition to other problems, the ratings must be discussed between the rating supervisor and the employee and can have important consequences for the employee. These ratings were, it seems, used in the early studies because of their availability, in the anticipation that other measures could be developed and administered in due course.

(3) Experimental Ratings.—By review of existing job descriptions, by interviews to determine "critical incidents" of the jobs, and by technical assistance and consultation with advisory committees consisting of representative incumbents and supervisory personnel, new "experimental" rating forms were developed for use in the Farrar-McLaurin studies. The forms consist of twelve rating categories for each of the two jobs, the categories relating to personality characteristics, job knowledge, and abilities found through the process to be relevant to job per-

formance. Each is rated on a seven-point scale (poor = 1 to outstanding = 7), with 3 being fixed as adequate. For the police form, weights were developed by the advisory committee to indicate relative importance of the categories to overall job performance.

Raters—the employees' supervisors—were given, in person and in writing, standardized instructions for use of the forms. To prevent the "halo" effect, supervisors rated all their subordinates on one category before proceeding to rate them on the next, etc. The raters were told that their evaluations were confidential and would not be used for any purpose other than the evaluation of the tests.

The court is impressed that the experimental rating method so developed represents an appropriate criterion measure for the jobs in question. These jobs are not ones which lend themselves to some objective measure, such as the sales produced by a sales representative. The principal limitations with the ratings so obtained are the lack of evidence as to reliability and the lack of special steps to detect or control possible bias.

Study Findings

Key findings from the Farrar-McLaurin studies are tabulated below. Correlations are shown only where presented in the body or exhibits of their reports. Statistically significant differences in means between two sub-groups (blacks and whites; short-tenure and long-tenure firefighters) are indicated by so de..gnating the lower of the two means.

	test	academy ave		efficiency		rating	experimental		rating
	mean	mean	r	mean	r	acad r	mean	r	acad r
Police									
Applicants									
T = 479	65.70								
W = 303	75.21								
B = 176	49.33**								
Officers									
T = 109	82.20		.46**		.12				
Officers									
T = 139	69.35								
W = 101	75.03								
B = 38	54.24**								
Officers									
T = 59	81.00	88.66	.77**	81.73	.31**				
Officers									
T = 96	80.46	87.94	.72**	81.10	.20	.19			
W = 76	84.54	89.00	.64**	81.44	.05	.08			
B = 20	64.95**	83.90**	.46*	79.81*	.17	−.05			
Officers									
T = 148	82.35	87.38	.45**	84.06	.08		433.79	.21*	
W = 140	83.11	87.58	.42**	84.11			434.75		
B = 8	69.12**	83.90**		83.18			417.00		
(B = 49)	66.78**	83.82**	.47**						
Deputies									
T = 83	82.58	87.36	.72**						
Officers									
T = 77	79.71	87.50	.60**						
Firefighters									
Applicants									
T = 507	70.99								
Employees									
T = 162	83.90	90.60	.43**	78.27	.12	−.23**			
Employees									
T = 196	81.99	91.27	.45**	79.45	.20**	−.06	61.24	.08	−.09
ST = 103	81.01	91.98	.37**	77.14**	.24**	.11	58.65**	.25**	.21**
LT = 93	83.09	90.47*	.62**	82.02	.05	−.01	64.11	−.20*	−.20

* $p < .05$

** $p < .01$

Fairness and Differential Validity

If members of one racial group generally obtain lower test scores than members of another group and those differences are not reflected in differences in measures of job performance, there is a need, where technically feasible, to investigate for possible unfairness of the test to the first group. See DOJ Guidelines, § 12b(7) (noting that this need increases the greater the severity of the adverse impact on the lower-scoring group). As the tabulation indicates, the Farrar-McLaurin studies do show that blacks as a group have scored lower on the 10—C than have whites. Indeed, in each of their studies where those scores are reported separately for the two racial groups, the differences are significant as $p < .01$.

One possibility is that the predictive validity of the test for one racial group is significantly different than for the other group. The inquiry here, as emphasized in A.P.A. Standard E9, is not whether there are differences in the correlation coefficients or whether one coefficient is statistically significant while the other is not. Rather, the proper statistical procedure is to test for significant differences in the coefficients.

In the one study in which a sufficient[26] number of both blacks and whites are involved, Drs. Farrar and McLaurin have performed such an analysis. Using the report of test scores and academy scores for 140 whites and 49 black police officers, they tested the correlation coefficients (where the coefficient for either subgroup was significant at $p < .05$) after z transformations, for significance of difference. None was significant at $p < .05$, and with respect to only one course (accident investigation) was the coefficient significant even at $p < .10$.

Failure, however, to reject the hypothesis that the correlation coefficients are the same for both groups is not by itself sufficient to demonstrate fairness. Where, as in the present case, test scores by two groups are used in the same manner for members of both groups, it is on the assumption that in general an in-

dividual's test score will appropriately predict his standing on the criterion whether he is a member of one group or the other. The predictive relationship between the test score and the criterion can be represented by a regression line[27] formula for converting a test score into a predicted criterion score. While regression lines can be calculated separately for the two groups and will almost always be somewhat different, it is important to know whether the slopes or intercepts (or both) of the lines are sufficiently different to call for abandonment of a common line for two groups. Otherwise, the common regression line (which is the effect of using test scores in the same way for both groups) may systematically underpredict for members of one group (while overpredicting for the other) their criterion score from a particular test score. The method for this inquiry, called analysis of variance, involves use of the *F* distribution tables for statistical significance. If desired, one can determine for what test scores the common regression line should, and should not, be abandoned.

Significantly, different regression lines may have the same or similar correlation coefficients. In such a situation comparison of the coefficients will not reveal the inappropriateness of using a common regression line. Thus, with the sample containing 76 white and 20[28] black police officers, comparison of the coefficients respecting 10—C scores (and modified 10—C scores) and either academy averages or efficiency ratings does not, as to any comparison, lead to rejection of the hypothesis of the coefficients being the same. Yet, when the same data are reviewed by analysis of covariance, as the court has done[29] for the hypothesis that a common regression line fits both whites and blacks, the obtained *F* ratios are 18.38 (10—C and academy average), 3.97 (10—C and efficiency rating), 26.91 (modified 10—C and academy average), and 4.18 (modified 10—C and efficiency rating), each of which (with $n_1 = 1$ and $n = 93$) is significant at $p < .05$. It is interesting that in the development of the modified 80-item scoring key for the 10—C, adopted to increase

the validity coefficient for blacks, the evidence becomes stronger that a common regression line should not be used for both groups.

In view of the fact that covariance analysis suggests rejection of a common regression line for both racial groups, one is tempted—since the differences between white and black means appear to be far greater on test scores than on the criterion-measures—to conclude that the performance of blacks is being underpredicted by the 10—C. However, if regression lines are computed separately for blacks and whites using the results of any of the studies where their test scores and criterion scores are reported separately, it will be found that the lines cross and that for test scores below that crossing point the criterion scores thereby predicted for blacks are less than for whites for the same test scores. Above that point there would be underprediction for blacks, but the intersections occur at such high test scores (the lowest point from any of the data is at a raw test score of 87) that few blacks would actually be affected. If one looks to see where any overprediction or underprediction is statistically significant, it is found that the only significant range of scores is for lower scores, where blacks are being overpredicted by the 10—C.[30]

The net result is that use of 10—C test scores in the same manner for both blacks and whites does not appear to be underpredicting the performance of blacks at the academy or on efficiency ratings. This analysis does not, of course, deal with the possibility of bias affecting the scores blacks obtain at the academy or on efficiency ratings; it only serves as a foundation for concluding that the 10—C is not to be criticized on the basis of differential validity inquiries. The 20—B cannot be subjected to these inquiries at the present time for lack of sufficient blacks in any study group.

Operational Utility

It is not, however, sufficient that, as here, a test is shown to have a statistically significant relationship to one or more criteria and not to be differentially unfair to the adversely affected racial group. In addition, in words of the EEOC Guidelines, § 1607.5(c), the relationship between the test and the criterion must have "practical significance" or, in the words of the DOJ Guidelines, § 12b(5), the usage of the test must be evaluated "to assure that it is appropriate for operational use." With different words, the two Guidelines are raising the same concern.

In concluding that the 10—C and 20—B tests are valid screening instruments, Drs. Farrar and McLaurin have emphasized their significant relationship to grades in the training academies, and this relationship cannot be doubted. However, as already indicated, it is the court's conclusion that *relative* standing in the academies, as distinguished from successful completion of academy training, is not an appropriate criterion unless it also be demonstrated that those academy grades are themselves valid predictors of job performance. The studies reflect, however, that for the most part the correlation between academy grades and measures of job performance are not significant and, in the few instances where significant correlations are found, the findings are mixed—some being positive and other being negative. A negative correlation, of course, indicates that the higher the academy grades, the lower the performance ratings tend to be. Although an employer is permitted to select the best person for the job despite resulting impact on a racial group, it is not permitted to engage in such selection procedures merely to employ the best person for training.

Nor has it here been demonstrated that either test is a valid predictor of successful completion of the required training courses. According to the director of the policy academy, only 11 of the 733 cadets attending the academy since 1962 have

failed to complete the training because of inadequate grades—and no data has been presented as to their 10—C test scores. According to the director of the firefighters academy, no student has failed because of poor grades, of course, since historical use of screening tests (whether the present ones or their predecessors) has imposed a restriction of range, the conclusion does not necessarily follow that every applicant could complete the training no matter how low his 10—C or 20—B score. However, it should be noted that on ocassion the raw test scores used to determine hiring eligibility have been substantially reduced, without apparent impact on their successful completion of the academies. And while recognizing that, as noted by Dr. McLaurin, this is not the regression lines developed from any of the studies would predict passing academy averages even for persons scoring zero on the 10—C and 20—B tests.

As earlier discussed, the regular efficiency ratings are not trustworthy criterion-measures of actual job performance. Even if they were, the studies provide inconclusive findings with respect to the 20—B (a significant correlation with a group of 196 firefighters, though only of a magnitude of .20, and an insignificant correlation of .12 with a group of 162 firefighters), and even more dubious results with respect to the 10—C.[32]

The experimental ratings are, as previously indicated, considered by the court as an appropriate criterion measure. The correlation, however, between the 20—B and these ratings is found to be .08, which is, of course, not significant; and, while a significant positive correlation is found with respect to the 103 firefighters having less than 3 years service, a significant negative correlation is found for the 93 having at least 3 years of tenure. Presumably, higher scores on the 20—B (which carried over to higher scores during academy training) resulted in better job performance for the first few years. Had this advantage merely been erased after more time on the job, this would be one matter—but, as stated, the findings actually showed a significant negative correlation for the longer tenured fire-

fighters, suggesting that over time the lower scoring applicants made the better employees. Absent any indication that during the first few years the lower scoring applicants had been inadequate on the job, one is hard pressed to conclude that the higher scoring 20—B applicants are in fact the better persons to hire. Further study might, of course, lead to other interpretations, such as a determination that recent improvements in the training given at the firefighters academy will result in better employees not only initially but also over time—but no such conclusions can be supported on the present evidence.

The correlation between 10—C and the experimental ratings is, with 148 in the sample, significant at $p < .05$, but has a magnitude of only .21. What do these figures mean? To begin with, it should be understood that for a correlation coefficient to be found significant at $p < .05$ is equivalent to saying that, if in fact no relationship between the two variables exists for the "population", the obtained results could be expected to occur only once in twenty such samples—and that therefore one can be 95% confident that for the population (of applicants) there is *some* correlation (or relationship) between the two variables. It does not mean that one can be 95% confident that the population coefficient is .21. Indeed, to state the population coefficient with only a 5% chance of error (*i.e.*, $\pi < .05$) requires use of a confidence *interval:* here, with a sample of 148, that the true coefficient lies somewhere between .0504 and .3592. The coefficient obtained from the sample is but an estimate of that true population coefficient.

A second consideration is to look at the test scores in the particular sample in comparison with the scores of all persons in the population. Where, as here, there is a restriction in the range of test scores of those in the sample because of prior use of the test, a statistical technique, called correction for restriction of range, may be appropriate for determining the magnitude of the correlation. This "corrected" coefficient, as reported by Drs. Farrar and McLaurin, is .36. It may be noted that utilization of the

correlation involves the assumption that the two variables (test scores and experimental ratings) are for the total population "normally distributed;" and, insofar as rating scores are concerned, it is just that—an assumption. Hence, it involves the same type of risk as does the use of a regression formula for values beyond the sample on which based—a technique which, during the trial, provoked Dr. McLaurin's criticism.

In general, other factors remaining the same, the greater the magnitude of the coefficient the more likely it is that the test will be appropriate for use. See DOJ Guidelines, §12b(5). The importance of the size of the correlation coefficient can perhaps best be understood by reference to certain basic statistical concepts. The square of the correlation coefficient, called the "coefficient of determination," gives the proportion of the variance of the criterion scores which is accountable by reference to variance of scores on the predictor test. Thus, with a correlation coefficient of .21, the study indicates that 4.4% of the variance among experimental ratings is explainable by reference to the variance in test scores, while 95.6% is not. Using the "corrected" coefficient of .36, still only 13% of the variance among experimental ratings could be accounted for by test score variance. By another formula, the correlation coefficient can be converted into a "coefficient of alienation", which gives the size of the error in attempting to predict experimental rating scores from test scores *relative* to the error that would result from a mere guess, *i.e.*, by not using the test. This calculation, based on a correlation coefficient of .21, reflects that use of the test predicts experimental rating scores with a margin of error that is only 2% smaller than it would be without the test, and, if based on the "corrected" coefficient of .36, indicates that the margin of error is only 6.7% less than what would occur by mere guess.

Anastasi comments, and quite properly so, that evaluation of a test in terms of the error of estimate will for many testing purposes be unrealistically stringent. Anastasi, PSYCHOLOGICAL

TESTING, p. 166 (4th Ed. 1976).[33] She notes that even tests with an unusually high validity of .80 would appear to be inefficient if used to predict individuals' relative standing on some criterion, but that most tests are merely used to determine which individuals will exceed a given minimum standard of performance or cutoff point in the criterion. The 10—C, of course, is utilized here both to screen applicants (cutoff scores) and to rank those passing applicants.

While the magnitude of the correlation coefficient is obviously of great importance, there is no minimum coefficient applicable to all employment situations. See DOJ Guidelines § 12b(5). "Under certain circumstances, even validities as low as .20 or .30 may justify inclusion of the test in a selection program." Anastasi, *op. cit.*, p. 166.

Another approach towards evaluation of the relationship found is to investigate the meaning of differences in test scores in relation to the differences in criterion scores thereby predicted. This involves use of the regression formula, which can be calculated from the correlation coefficient and the means and standard deviations of the two variables. Thus, Dr. Roland Ramsey, the plaintiffs' expert witness, questioned the practical value of the 10—C by noting, from the Farrar-McLaurin study involving 140 whites and 49 blacks, that an increase in raw test scores of 40 points produced, under the regression lines given, less than 5 points increase in predicted academy averages.

The common regression line computed for the 148 officers with both test scores and experimental ratings is $\bar{y} = 300 + .162x$, where x represents a given test score and $\bar{y}$ is the rating predicted thereby. At first glance, this regression line does not appear to be subject to the criticism made by Dr. Ramsey respecting the other study, for it will be seen that, for example, a test score difference of 10 will predict an experimental rating difference at 16. However, it should be understood that the linear regression formula (as well as the criterion

mean and criterion standard deviation, although not the correlation coefficient) varies in direct proportion to any factor by which criterion scores in the sample have been multiplied. The Farrar-McLaurin study reports the overall experimental rating as the summation of weighted components which comprise the rating. For example, a sample subject rated as 5 (very good) on each of the 12 rating components would, because of the method, be reported as having a rating score of 499, while another subject identically rated except for scores of 4 (good) on the appearance and dependability components would receive an overall rating of 483, with 698 representing a "perfect" score of 7's on all components.

To prevent potential misinterpretation, it is well to consider the regression line not only in the form in which expressed in the Farrar-McLaurin studies, but also in a form which does not contain the inflation caused by the weighting procedure. This can be done, while still retaining the concept of the components having differing weights, by expressing the weights in a manner in which the average weight is 1. That is, instead of the "communication" component having a weight of 8.78 (as reported in the study), of "problem solving" a weight of 9.44, of "learning" a weight of 8.22, etc., they can be shown as having weights of 1.056, 1.136, and .989, etc., respectively. Then, a summation of weighted component scores for a "perfect" score of 7 on all 12 components would result in 84, identical with a "perfect" score if not weighted. Not only does this method retain the concept of weighing the different components, but the transformation (whether of means, standard deviations, or regression line) can simply be made by dividing the reported results by a constant, here 8.31333. The obtained regression line is $\bar{y} = 36.087 + 195x$, where x represents the test score and $\bar{y}$ is the predicted rating (using the new method of expressing weights). It will now be seen that, as with the other studies criticized by Dr. Ramsey, a large difference in test scores produces only a small difference in predicted (un-inflated) ex-

perimental rating scores, *e.g.*, a 40 point raw score difference on the test gives less than 8 points difference on the rating score.

Another method for evaluation, which is not complicated by the weighting procedure, is to consider the "standard error of estimate", which, for the data analyzed, is computed to be 92.63. Use of this statistic is demonstrated as follows: while a raw test score of 70 through the regression formula predicts an experimental rating of 413, one can through use of the standard error of estimate determine, at $p < .05$, the experimental rating actually to lie within the range of 231 to 595. Similarly, the predicted experimental rating, at $p. < .05$, from a raw test score of 40 is found to be in the range of 183 to 547. Obviously, there is a potential overlap, where persons with raw scores of 40 and 70 on the test may nevertheless obtain the same experimental rating score. It is, furthermore, possible to determine how much of a difference in test scores is required for one to be able to predict, at $p < .05$, that the higher-scoring applicant will receive an experimental rating which also is higher[34]—and this calculation results in a finding that a difference in test scores of over 86 raw points is necessary for such a conclusion to be reached. It should be noted that the total range of raw test scores to this date used to rank successful applicants (*i.e.*, from a low raw score of 48 to the perfect score of 120, which has not been obtained by any) is yet too limited to enable one to say, at the .05 level, that the highest-scoring applicant would be predicted to obtain a higher experimental rating than the lowest-scoring applicant.

Since the 10—C is utilized not only in an attempt to rank the successful candidates, but also to screen the unsuccessful, it is appropriate to analyze the study results with respect to minimum experimental ratings and to predictions for persons scoring at, and below, the test cut-off scores. A test score of 48, the lowest used as a cut-off, yields a predicted experimental rating of 378, or an average unweighted rating on each component of the experimental rating of 3.78 (3 = adequate,

4 = good).[35]. Common regression lines can, of course, also be computed separately for each of the twelve components of the rating. When this is done, one finds that a test score of 48 will as to each component predict an unweighted rating of 3 or above, *i.e.*, at least "adequate." While recognizing the risk in extrapolation beyond the range of sample scores,[36] but having little else available for comparable analysis, one can look to estimates of experimental ratings predicted by the regression lines for test scores below 48. If this is done, it appears that even with a test score of 0 the predicted rating is "adequate" or above for the rating as a whole and for seven of the twelve components. Even as to the five components for which the estimated unweighted experimental rating from a 0 test score would be less than 3, the predicted rating cannot be said to be less than "adequate" at $p < .05$.

A technique for evaluating tests which employ cut-off scores for screening purposes is to consider "false positives" (persons scoring below the cutoff but nevertheless scoring above the acceptable level of performance on the criterion) in relation to "false acceptances" (persons scoring above the cutoff but below the acceptable level of performance), thereby leading to a comparison between the relative percentage of successful employees above and below the cut-off scores. However, neither this method nor the Taylor-Russell tables (used to estimate net gain in selection accuracy through test usage) can be directly used in the present case because all employees for whom "success" data are available have been screened by the test. It is possible to project the "base rate" (through use of the regression line, standard error of estimate, and normal distribution curves) and then to conduct such inquiries, and, if this be done, one finds any incremental validity to be negligible.

Still another approach is to estimate the effect of the test not on the percentage of persons exceeding minimum performance, but on overall performance of the selected persons. A table

given by Anastasi, *op. cit.* at p. 173, gives the expected rise in criterion scores through test usage in relation to its validity coefficient and the selection ratio. In the present case, with a .21 coefficient and a selection rate of .19 (506 officers hired from 2721 applicants), and with a standard deviation of known criterion scores of 94.74, one finds from the table that use of the test (had the applicants actually been hired in the order of the test scores) would probably have produced an average gain of 27 points in the total weighted experimental rating (over the rating expected had the test not been used). This gain is equivalent to being rated one point higher on three of the 12 rating components. The average unweighted rating on each of the components would, without the test, have been 4.07 (4 = good), which compares to 4.35, using the test.

The assessment of utility of a test which, like the 10—C, has a statistically significant validity, albeit of very low magnitude, must include certain value judgments. One of these involves consideration of the nature of the job in question and the consequences of a faulty hiring decision. There can be little dispute that police officers perform a vital, and sensitive, function in our society. The desirability of "upgrading" of law enforcement has been emphasized in two reports received in evidence, the 1967 Task Force Report on the Police, issued by the President's Commission on Law Enforcement and the Administration of Justice, and the 1973 Report on Police, issued by the National Advisory Commission on Criminal Justice Standards and Goals. Economic costs are also involved, particularly in view of the cost of academy training of officers and the restraints placed upon discharge of marginal officers under the civil service laws.

Without demeaning the importance of law enforcement officials, however, it can hardly be said that the possibility of occasional selection of an inept officer presents the same type of daily economic and human risk factors as is involved, for example, in the employment of airline pilots or bus-drivers. *Cf. Spurlock* v. *United Airlines, Inc.*, [5 EPD ¶ 7996] 475 F.2d 216,

219 (CA7 1972); *Usery* v. *Tamiami Trail Tours, Inc.*, [11 EPD ¶10,916] 531 F.2d 224 (CA5 1976) (Age Discrimination in Employment Act case, with somewhat related question). A twelve months' probationary period is provided, during which time the occasional incompetent may be detected and dismissed and during part of which time the employee is undergoing training rather than being "on the street". The principal public concern, it would appear, is not so much that the most able officers be employed (though that certainly, would be desirable) as that the emotionally unfit not be employed. In this context, it is perhaps noteworthy that the 1967 Presidential Commission's report contained a recommendation for use of psychological tests to detect applicants with personality defects (but no such recommendation respecting aptitude tests); and the 1973 National Advisory Commission's report, while acknowledging the desirability of valid aptitude tests, was skeptical as to the results of research to that date. So far as the court has been informed, the 10—C was not designed, and has not been validated, for use in detecting emotional disorders or defects.

The DOJ Guidelines § 12b(b), provide that, in determining operational appropriateness, one should consider "the degree of adverse impact of the procedure, the availability of other selection procedures of greater or substantially equal validity, and the need of an employer, required by law or regulation to follow merit principles, to have an objective system of selection." Obviously this latter factor (requirement under civil service law to give some objective test) cannot by itself suffice as justification for a test which has, as here, substantial adverse impact on a racial group. While it can be said that no other available" selection with greater validity than the 10—C has been found, yet it must also be said—considering the minimal benefits resulting from the 10—C in the context of this employment situation—that no-test-at-all has "substantially" the same practical validity as the 10—C.

In summary, the 20—B Firefighter test has not been shown to be a valid predictor of a job-relevant criterion measure and the

10—B Policeman test, while having a statistically significant relationship of a very low magnitude with a job-relevant criterion measure, has not been shown to be appropriate for operational use in screening or ranking applicants."

Violation and Remedy

Having concluded that use of the 10—C and 20—B has had an adverse impact upon black applicants and that the studies presented fail to demonstrate job-relatedness, the court must nevertheless determine when the requirements of law were violated and what relief is appropriate therefor. This inquiry should involve no less care then consideration of the tests themselves.

The requirements of Title VII first became applicable to the Personnel Board in March 1972. At that time, and for many years earlier, the Board was required by state law to administer appropriate tests to screen and rank applicants—a requirement which continues to the present time, subject to any over-riding proscriptions of Title VII. It had several years earlier selection the 10—C and 20—B tests as the best tests then available, with the hope that black applicants would fare better than under previous tests. By March 1972 a preliminary, in-house validity study had been conducted, which reflected some improvement in hiring of blacks and the indication of appropriate validity based upon relationship with existing criterion measures. An in-depth independent validation study was immediately undertaken, including investigation of alternative or supplemental selection procedures to improve the predictive validity or decrease adverse impact upon blacks. At least since 1965 the Board has not intentionally discriminated against black applicants but, to the contrary, has attempted to increase black employment within the options available under state law, including modification of the scoring key for the 10—C when recommended by the consultants as a method for increasing validity of the test for black applicants.

The preliminary reports from the consultants, made while more trustworthy measures of job performance were being developed, contained signs of potential validity and recommended continued usage of the test pending the additional studies. Not until April 25, 1975, with respect to the 10—C, and July 8, 1976, with respect to the 20—B, were the studies using these new criterion measures completed and reported to the Board. It was on these respective dates that, in the court's opinion, it should have been concluded that provisional use of the tests was no longer permissible. Prior thereto, thc Board was, in the court's opinion, justified in continuing to use the tests (and the eligibility lists generated therefrom) in anticipation of favorable results from those studies. Use of the tests (or of the eligibility lists therefrom) was thereafter, however, contrary to the requirements of Title VII, which override state law inconsistent therewith.

The remedy should be appropriate to the violation found. In this case, from X-11, it is found that, for the two administrations of the 10—C from which eligibility lists used after April 25, 1975, were formed, 658 (or 88%) of the 747 white applicants were placed on the eligibility lists. Had a like percentage of black applicants been so placed, a total of 252 would have been on the lists—128 more than actually placed on the list. Accordingly, to the extent they are still interested, an additional 128 blacks from the prior administrations of the test should be added to the present eligibility lists. This remedy only relates to prohibited use of the 10—C as a screening instrument. An additional measure is needed to correct for the improper use of the test as ranking procedure. Had the eligibility lists been representative of the applicant group and had certifications from the list likewise been representative of the racial composition of the list, approximately 28% of the persons certified would have been black. It is clear that there has been "under-certification" of blacks by this standard, although the precise degree cannot be determined from evidence before the court, which gives such information only by calendar years. The Board is directed to

ascertain the extent of such under-certification and in future certifications to include at least 1 black candidate for every 3 certified until such time that, considering the certifications after April 25, 1975, total number of blacks certified becomes 28% of the total number of persons certified. Thereafter (and until some new selection procedures are adopted which are sufficiently job-related or which have no adverse impact upon blacks) at least 2 of every 7 persons certified by the Board from the revised present list shall be black, provided there be a sufficient number of black applicants interested.

A similar investigation of X-11 with respect to the 20—B, where only one eligibility list has been in effect since July 8, 1976 (the date of the report involving the experimental ratings), results in a conclusion that 91 black applicants should be added to the present eligibility list for firefighters, that at least 1 of every 3 persons hereafter certified shall be black until such time that (considering certifications after July 8, 1976) the total number of blacks certified becomes 14%[19] of the total number of persons certified, and that thereafter (pending adoption of some other valid or nondiscriminatory selection instrument) at least 1 of every 7 certified by the Board from the revised current list shall be black.

This order does not preclude use of the 10—C or 20—B as a device for ranking one white as against another white, or one black as against another black. Such a use may be made by the Board, if it so desires, without any discriminatory impact on a racial group. The order does not prevent the Board from new administrations of the 10—C or 20—B (or other tests) or from forming new eligibility lists from time to time; provided, however, that, unless and until a selection instrument is found which either has no adverse impact racially or is sufficiently valid, the test results shall be used in a manner consistent with this opinion, *i.e.*, the eligibility list and certifications to be representative racially of the applicant group regardless of test scores.

APPENDIX C

Footnotes

[1] Fourteen separate county and municipal employers are covered by the law. Cities with a population of under 5,000 are excluded.

[2] With multiple vacancies, the number of persons certified is two more than the number of vacancies to be filled.

[3] Not presently at issue are requirements (such as age or education) which may be imposed as conditions to taking an examination, nor are specifications (such as residence within Jefferson County) which may give preference to certain applicants.

[4] Unless otherwise noted, reference to police officers in the balance of this opinion will also refer to deputy sheriffs.

[5] Under F.R.Civ.P. Rule 42, the four actions were consolidated with respect to challenges to Personnel Board tests and a separate trial was scheduled respecting the attacks on the Policeman 10—C, Firefighter 20—B, and Office Worker 30—B tests. At the trial the plaintiffs indicated that the attack on the Office Worker 30—B test was dropped for lack of evidence of adverse impact and that any attack on the 10—C and 20—B tests based on sex was likewise dropped for lack of evidence.

[6] The 10—C and 20—B tests were adopted by the Board after initially experimenting, commencing in January 1966, with alternate forms of the IPMA tests.

[7] The Board's studies resulted in selection of the 10—C form because of its significant and positive correlation with a greater number of the selected criteria measures than did the alternate IPMA form. The study indicated a significant and positive correlation between 10—C scores and training academy average (as well as several course grades in the academy) and between the training academy average and the officers' latest efficiency ratings.

[8] Not until 1968 was the residency requirement of the City of Birmingham removed by the city ordinance. A Jefferson County preference remains in effect, but this can hardly disadvantage blacks, who constitute a larger portion of the Jefferson County population than of neighboring counties.

[9] An intent to discriminate would presumably be required for there to be a violation of 42 U.S.C. § 1983, if not of 42 U.S.C. § 1981. See *Washington* v. *Davis*,.—U.S.—(June 7, 1976).

[10] *Albemarle Paper Co.* v. *Moody*, [9 EPD ¶ 10,230] 422 U.S. 405, 425 (1975).

[11] *Griggs* v. *Duke Power Co.*, [3 EPD ¶8137] 401 U.S. 424, 432 (1971).

[12] There are some conflicts between the EEOC and the DOJ Guidelines. However, it is not necessary as to the issues presently before the court that a choice be made between the two.

[13] Information for this table has been taken from X-11 and from data respecting hires supplied by the parties at the court's request following formal close of the evidence. Certain caveats should be noted: The results of the 10—C exam administered on April 29-30, 1971, have been eliminated because it was not used for employment decisions after March 24, 1972. The results of the 10—C exam administered on September 30 and October 1, 1971 and of the 20—B exam administered on May 26-27, 1971, have been included in the tabulation, even though in part the eligibility lists taken therefrom would have been used prior to March 24, 1972. The number of hires includes those hired in 1972 prior to March 24, 1972. As the Board points out, the number of hires is affected by voluntary choices of the candidates (such as declining job offers or waiving consideration), but, lacking reliable data on such matters for both whites and blacks, the court has looked to actual hires as the measure of the overall selection ratios. Finally, it should be noted that, since persons are permitted to take exams more than once, the applicant figures do not completely accurately reflect the number of different individuals involved. These limitations do not, in the court's opinion, prevent meaningful usage of the data for the purposes indicated.

[14] These figures are derived from X-12 and are subject to the appropriate caveats indicated in fn. 13, *supra*. Moreover, X-12 does not have any information for one eligibility list and does not contain percentile information as to several lists. In a very real sense, "passing" an exam is measured not by obtaining a derived score of at least 70 (and thereby being entered on the eligibility list), but by obtaining a score sufficiently high to be placed on the eligibility list at a position where, during use of that list, the candidate will actually be certified to an employing agency.

[15] Until a couple of months prior to trial, the litigation was being prepared with the anticipation that all tests under challenge were to be considered at a single hearing. When the decision was made by the court that the first trial would only concern the 10—C, 20—B, and 30—B tests, counsel and witnesses were freed to shift their attention to "loose-ends" on these three tests.

[16] Some items on each test elicit knowledge which apparently would be needed for performance of job functions; others do not. Some effort has been made by the test developer to give "face validity", as by expressing an item involving numerical problem solving in the context of information with which job occupants would be dealing. Face validity does not affect validity for usage as a selection procedure so much as it may overcome motivational resistance by those taking the test.

[17] "Reliability refers to the consistency of scores obtained by the same persons when reexamined on the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions." Anastasi, PSYCHOLOGICAL TESTING, p. 103 (4th Ed. 1976). The methods used in this study to estimate reliability were the split-half technique (corrected by the Spearman-Brown prophecy formula) and the Kuder-Richardson formula 20.

[18] A test item correctly answered by too high a proportion of the applicants is considered not difficult enough; correctly answered by too low a proportion, it is considered too difficult. In this study items correctly answered by 30 to 70% of the applicants were considered satisfactorily difficult.

[19] Item "discrimination" is in essence a comparison between scores made on an individual test item and scores made on the total test, thereby ascertaining whether particular items "discriminate" significantly in predicting success on the test as a whole. In the study, significance was established at $p < .05$.

[20] As previously indicated, the 10—C exam was in fact modified through use, effective April 10, 1974, of an 80-item answer key, which had the effect of eliminating for scoring purposes 40 of the items. This particular change was done to improve correlation with an academy average criterion for blacks (rather than to improve item difficulty or discrimination levels), but it indicates the search for increased test utility. The consultants also recommended consideration of possible modification (or supplementation) of 20—B to improve levels of item difficulty and discrimination. Only one administration of the 20—B was given after this recommendation, and that was done when the existing eligibility list had almost been exhausted but mass administration of an additional test (the PAS) had not yet become feasible.

[21] Neither the EEOC nor the DOJ Guidelines require reporting of raw data statistics for Σx, Σx^2, Σy, Σy^2, or Σxy. Such information would, however, be helpful, permitting application of statistical measures not chosen by authors of the report without the loss of accuracy which results from derivation of such items from means, standard deviations and correlation coefficients.

[22] The study of the sample 59 police officers for whom Raven and PAS test scores were available is subject to question for voluntarism. See A.P.A. Standard E6.1.2.

[23] The possibility of bias is of particular concern where subjective evaluations are used as criteria and there are significant differences in those measures for different racial groups. See DOJ § 12b(2). In each study where means and standard deviations are presented separately for blacks and whites with respect to one or more of the basic criterion-measures (academy average, efficiency rating, or experimental rating), the means for blacks is less than for whites, and only in the study which involved but 8 blacks were any of these differences not statistically significant at least at $p < .05$. It may be argued that the academy grades should be treated as objective (being based in major part upon multiple-choice exams), but a substantial part of those grades is apparently dependent upon instructors' subjective appraisal of students' performance. (This latter comment is not intended to suggest that more paper-and-pencil tests should be used in the academies, but rather that even academy grades are at least in part subjective measures.)

[24] See fn. 17, *supra*. Inquiry into the reliability of a criterion is not a trivial consideration. See A.P.A. Standards E4.4. That the various techniques for estimating such reliability have their own limitations affects the interpretation to be reached, not the desirability of making the effort.

[25] Of course, relative standing or grades in the academy may, depending upon the statistical technique employed, be of use in correlating test scores with successful completion of the training.

[26] The DOJ Guidelines, § 12b(7)(v)(1), do not require analysis where less than thirty persons are in either of the subgroups.

[27] Scattergrams in the present case do indicate, within the ranges of scores available, that relationships between the predictors and criteria for both racial groups are essentially linear.

[28] One should view with caution differential studies where either group has less than 30 members. However, the principal concern is that true differences will not appear to be significant with smaller sample numbers.

[29] The court has analyzed the 96-subject sample rather than the 189-subject sample because the report of the former includes deviation and correlation data not only for the two racial groups but also for the sample as a whole, facilitating analysis. It should be recognized that analysis from such statistics are subject to rounding errors which could have been avoided had raw data summations been provided.

[30] Absent "adverse impact" on whites as a whole, the "under-predicted" whites cannot challenge the test under Title VII.

[31] Regression lines should not generally be used for prediction based on predictor scores which are beyond the range of predictor scores found in the sample, for beyond such known scores the relationship may cease to be significant, may cease to be linear, or may have a slope change. Of course the same possibilities exist when one attempts to justify non-selection based upon correlation coefficients developed through the subjects selected or when one attempts to modify coefficients for restriction of range, such as Drs. Farrar and McLaurin have done.

[32] The correlations between the 10—C and efficiency ratings were significant only with respect to the volunteer group of 59 officers.

[33] The Anastasi volume was qualified during trial as a recognized treatise under F.R.E. 803(18). Additional standard texts used by the court for the purpose of taking judicial notice of basic statistical formulae are Guilford & Fruchter, Fundamental Statistics in Psychology and Education (5th Ed. 1973); Walker & Lev, Statistical Ineference (1953); Burrington & May, Handbook of Probability and Statistics (2nd Ed. 1970); and Mehrens & Lehamann, Standard Tests in Education (2nd Ed. 1975).

[34] A note should be made of this technique since not directly given in most texts. Using the normal curve, the possibility of scores exceeding $y/o = + .76$ is .2236 and likewise the possibility of scores being less than $y/o = + - .76$ is .2236. The possibility of both events occurring is $.2236^2$, or .05. Accordingly, there must be a separation of $2(.76) \times$ standard error of estimate for two predicted scores to be different at $p. < .05$. One can then determine the difference in predictor scores necessary to produce this separation in predicted scores.

[35] By comparison, a raw-test score of 106 (the highest reported in the study for any subject) yields a predicted experimental rating of 472, or an average unweighted rating on each component of 4.73 (5 = very good).

[36] See fn. 31, *supra,* and the discussion on page 21 of this opinion respecting assumption of normal distribution when a correlation coefficient is corrected for restriction of range.

[37] Studies of the PAS, developed by Drs. Farrar and McLaurin, show extremely high correlations with experimental ratings, as well as with academy training and efficiency ratings. Ironically, the results are so promising as to cause some concern as to a spurious relationship which may not be replicated. In any event, technical difficulties, unresolved to date, have prevented its administration on a wide-scale basis such as for all applicants, so that for practical purposes it is not "available".

[38] Correlation studies respecting efficiency ratings and "experimental" ratings were conducted only for officers employed by the City of Birmingham. However, the evidence is persuasive that job requirements for deputy sheriffs and for police officers employed by other municipalities are essentially the same as for Birmingham officers. The higher correlations found with respect to academy training are not of themselves sufficient to justify a conclusion as to operational validity for these officers different from that reached respecting Birmingham.

[39] Blacks constituted only 14% of the applicants on the only administration of the 20—B involved.

APPENDIX D

Ensley Branch of the N.A.A.C.P., Plaintiff v. George Seibels, et al., Defendants. Civil Action No. CA 74-Z-12-S.

John W. Martin et al., Plaintiffs v. City of Birmingham et al., Defendants. Civil Action No. CA 74-Z-17-S.

United States of America, Plaintiff v. Jefferson County et al., Defendants. Civil Action No. CA 75-P-0666-S.

Lucy Walker et al., Plaintiffs v. Jefferson County Home et al., Defendants. Civil Action No. CA 76-M-2047-S.

United States District Court, Northern District of Alabama, Southern Division. January 10, 1977.

ORDER

Pursuant to the findings and conclusions contained in the Memorandum of Opinion filed herewith, unavoidably protracted because of the need to detail the findings of fact and the reasons therefor, it is ordered as follows:

1. Use by the Personnel Board of Jefferson County of the 30-B Office Workers test has not violated Title VII or other applicable law.

2. Use by the Personnel Board of Jefferson County of the 10-C Policeman Test and the 20-B Firefighter Test has violated Title VII since April 25, 1975, and July 8, 1976, respectively.

3. To the current eligibility list for police officers and deputy sheriffs the Personnel Board shall add the names of 128 black applicants from prior administrations of the 10-C to the extent such number are still interested. In future certifications, at least 1 black on the revised eligibility list shall be certified for each 3 persons certified until such time that the total number of blacks certified after April 25, 1975, shall be 28% of the total number so certified. Thereafter during use of the current eligibility list as so revised, at least two persons of every seven certified shall be

black. Pending adoption of some selection procedure which either has no adverse effect upon black applicants or is sufficiently job-related, the number of blacks on any new eligibility list (and certified therefrom) shall be representative of the number of the black applicants.

4. To the current eligibility list for firefighters, the Personnel Board shall add the names of 91 black applicants from prior administrations of the 20-B to the extent such number are still interested. In future certifications, at least 1 black on the revised eligibility list shall be certified for each 3 persons certified until such time that the total number of blacks certified after July 8, 1976, shall be 14% of the total number so certified. Thereafter, during use of the current eligibility list as so revised, at least one person of every seven certified shall be black. Pending adoption of some selection procedure which either has no adverse impact upon black applicants or is sufficiently job-related, the number of blacks on any new eligibility list (and certified therefrom) shall be representative of the number of black applicants.

5. In accordance with F.R.Civ.P. Rule 55(b), the court expressly determines that there is no just reason for delay and expressly directs entry of judgment as to the issues here involved, namely, whether use by the Personnel Board of the 10-C, 20-B, and 30-B tests are proscribed by law and, if so, the appropriate remedy therefor.

APPENDIX E

Civil Rights Act of 1964, As Amended. 42 U.S.C. §2000(e)

DEFINITIONS

For the purposes of this title—

(a) The term "person" includes one or more individuals, governments, governmental agencies, political subdivisions, labor unions, partnerships, associations, corporations, legal representatives, mutual companies, joint-stock companies, trusts, unincorporated organizations, trustees, trustees in bankruptcy, or receivers.

(b) The term "employer" means a person engaged in an industry affecting commerce who has fifteen or more employees for each working day in each of twenty or more calendar weeks in the current or preceding calendar year, and any agent of such a person, but such term does not include (1) the United States, a corporation wholly owned by the Government of the United States, and Indian tribe, or any department or agency of the District of Columbia subject by statute to procedures of the competitive service (as defined in section 2102 of Title 5 of the United States Code), or (2) a bona fide private membership club (other than a labor organization) which is exempt from taxation under section 501(c) of the Internal Revenue Code of 1954, except that during the first year after the date of enactment of the Equal Employment Opportunity Act of 1972, persons having fewer than twenty-five employees (and their agents) shall not be considered employers.

(c) The term "employment agency" means any person regularly undertaking with or without compensation to procure employees for an employer or to procure for employees opportunities to work for an employer and includes an agent of such a person.

(d) The term "labor organization" means a labor organization engaged in an industry affecting commerce, and any agent of such an organization, and includes any organization of any kind, any agency, or employee representation committee, group, association, or plan so engaged in which employees participate and which exists for the purpose, in whole or in part, of dealing with employers concerning grievances, labor disputes, wages, rates of pay, hours, or other terms or conditions of employment, and any conference, general committee, joint or system board, or joint council so engaged which is subordinate to a national or international labor organization.

(e) A labor organization shall be deemed to be engaged in an industry affecting commerce if (1) it maintains or operates a hiring hall or hiring office which procures employees for an employer or procures for employees opportunities to work for an employer, or (2) the number of its members (or, where it is a labor organization composed of other labor organizations or their representatives, if the aggregate number of the members of such other labor organization) is (A) twenty-five or more during the first year after the date of enactment of the Equal Employment Opportunity Act of 1972, or (B) fifteen or more thereafter, and such labor organization—

(1) is the certified representative of employees under the provisions of the National Labor Relations Act, as amended, or the Railway Labor Act, as amended;

(2) although not certified, is a national or international labor organization or a local labor organization recognized or acting as the representative of employees of an employer or employers engaged in an industry affecting commerce; or

(3) has chartered a local labor organization or subsidiary body which is representing or actively seeking to represent employees of employers within the meaning of paragraph (1) or (2); or

(4) has been chartered by a labor organization representing or actively seeking to represent employees within the meaning of paragraph (1) or (2) as the local or subordinate body through which such employees may enjoy membership or become affiliated with such labor organization; or

(5) is a conference, general committee, joint or system board, or joint council subordinate to a national or international labor organization, which includes a labor organization engaged in an industry affecting commerce within the meaning of any of the preceding paragraphs of this subsection.

(f) The term "employee" means an individual employed by an employer, except that the term "employee" shall not include any person elected to public office in any State or political subdivision of any State by the qualified voters thereof, or any person chosen by such officer to be on such officer's personal staff, or an appointee on the policy making level or an immediate adviser with respect to the exercise of the constitutional or legal powers of the office. The exemption set forth in the preceding sentence shall not include employees subject to the civil service laws of a State government, governmental agency or political subdivision.

(g) The term "commerce" means trade, traffic, commerce, transportation, transmission, or communication among the several States; or between a State and any place outside thereof; or within the District of Columbia, or a possession of the United States; or between points in the same State but through a point outside thereof.

(h) The term "industry affecting commerce" means any activity, business, or industry in commerce or in which a labor dispute would hinder or obstruct commerce or the free flow of commerce and includes any activity or industry "affecting commerce" within the meaning of the Labor-Management Repor-

ting and Disclosure Act of 1959 and further includes any governmental industry, business, or activity.

(i) The term "State" includes a State of the United States, the District of Columbia, Puerto Rico, the Virgin Islands, American Samoa, Guam, Wake Island, the Canal Zone, and Outer Continental Shelf lands defined in the Outer Continental Shelf Lands Act.

(j) The term "religion" includes all aspects of religious observance and practice, as well as belief, unless an employer demonstrates that he is unable to reasonably accommodate to an employee's or prospective employee's religious observance or practice without undue hardship on the conduct of the employer's business.

(k) The terms "because of sex" or "on the basis of sex" include, but are not limited to, because of or on the basis of pregnancy, childbirth or related medical conditions; and women affected by pregnancy, childbirth, or related medical conditions shall be treated the same for all employment-related purposes, including receipt of benefits under fringe benefit programs, as other persons not so affected but similar in their ability or inability to work, and nothing in section 703(h) of this title shall be interpreted to permit otherwise. This subsection shall not require an employer to pay for health insurance benefits for abortion, except where the life of the mother would be endangered if the fetus were carried to term, or except where medical complications have arisen from an abortion: Provided, that nothing herein shall preclude an employer from providing abortion benefits or otherwise affect bargaining agreements in regard to abortion.

EXEMPTION

This title shall not apply to an employer with respect to the employment of aliens outside any State, or to a religious corporation, association, educational institution, or society with

respect to the employment of individuals of a particular religion to perform work connected with the carrying on by such corporation, association, educational institution, or society of its activities.

DISCRIMINATION BECAUSE OF RACE, COLOR, RELIGION, SEX, OR NATIONAL ORIGIN

(a) It shall be an unlawful employment practice for an employer—

(1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or

(2) to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.

(b) It shall be an unlawful employment practice for an employment agency to fail or refuse to refer for employment, or otherwise to discriminate against, any individual because of his race, color, religion, sex, or national origin, or to classify or refer for employment any individual on the basis of his race, color, religion, sex, or national origin.

(c) It shall be an unlawful employment practice for a labor organization—

(1) to exclude or to expel from its membership, or otherwise to discriminate against, any individual because of his race, color religion, sex, or national origin.

(2) to limit, segregate, or classify its membership or applicants for membership, or to classify of fail or refuse to refer for employment any individual, in any way which would deprive or tend to deprive any individual of employment opportunities, or would limit such employment opportunities or otherwise adversely affect his status as an employee or as an applicant for employment, because of such individual's race, color, religion, sex, or national origin; or

(3) to cause or attempt to cause an employer to discriminate against an individual in violation of this section.

(d) It shall be an unlawful employment practice for any employer, labor organization, or joint labor-management committee controlling apprenticeship or other training or retraining, including on-the-job training programs to discriminate against any individual because of his race, color, religion, sex, or national origin in admission to, or employment in, any program established to provide apprenticeship or other training.

(e) Notwithstanding any other provision of this title, (1) it shall not be an unlawful employment practice for an employer to hire and employ employees, for an employment agency to classify, or refer for employment any individual, for a labor organization to classify its membership or to classify or refer for employment any individual, or for an employer, labor organization, or joint labor-management committee controlling apprenticeship or other training or retraining programs to admit or employ any individual in any such program, on the basis of his religion, sex, or national origin in those certain instances where religion, sex, or national origin is a bona fide occupational qualification reasonably necessary to the normal operation of that particular business or enterprise, and (2) it shall not be an unlawful employment practice for a school, college, university, or other educational institution or institution of learning to hire and employ employees of a particular religion if such school,

college, university, or other educational institution or institution of learning is, in whole or in substantial part, owned, supported, controlled, or managed by a particular religion or by a particular religious corporation, association, or society, or if the curriculum of such school, college, university, or other educational institution or institution of learning is directed toward the propagation of a particular religion.

(f) As used in this title, the phrase "unlawful employment practice" shall not be deemed to include any action or measure taken by an employer, labor organization, joint labor-management committee, or employment agency with respect to an individual who is a member of the Communist Party of the United States or of any other organization required to register as a Communist-action or Communist-front organization by final order of the Subversive Activities Control Board pursuant to the Subversive Activities Control Act of 1950.

(g) Notwithstanding any other provision of this title, it shall not be an unlawful employment practice for an employer to fail or refuse to hire and employ andy individual for any position, for an employer to discharge any individual from any position, or for an employment agency to fail or refuse to refer any individual for employment in any position, or for a labor organization to fail or refuse to refer any individual for employment in any position, if—

(1) the occupancy of such position, or access to the premises in or upon which any part of the duties of such position is performed or is to be performed, is subject to any requirement imposed in the interest of the national security of the United States under any security program in effect pursuant to or administered under any statute of the United States or any Executive order the President; and

(2) such individual has not fulfilled or has ceased to fulfill that requirement.

(h) Notwithstanding any other provision of this title, it shall not be an unlawful employment practice for an employer to apply different standards of compensation, or different terms, conditions, or privileges of employment pursuant to a bona fide seniority or merit system, or a system which measures earnings by quantity or quality of production or to employees who work in different locations, provided that such differences are not the result of an intention to discriminate because of race, color, religion, sex, or national origin, nor shall it be an unlawful employment practice for an employer to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin. It shall not be an unlawful employment practice under this title for any employer to differentiate upon the basis of sex in determining the amount of the wages or compensation paid or to be paid to employees of such employer if such differentiation is authorized by the provisions of section 6(d) of the Fair Labor Standards Act of 1938, as amended (29 U. S. C. 206(d)).

(i) Nothing contained in this title shall apply to any business or enterprise on or near an Indian reservation with respect to any publicly announced employment practice of such business or enterprise under which a preferential treatment is given to any individual because he is an Indian living on or near a reservation.

(j) Nothing contained in this title shall be interpreted to require any employer, employment agency, labor organization, or joint labor-management committee subject to this title to grant preferential treatment to any individual or to any group because of the race, color, religion, sex, or national origin of such individual or group on account of an imbalance which may exist with respect to the total number or percentage of persons of any race, color, religion, sex, or national origin employed by any employer, referred or classified for employment by any employ-

ment agency or labor organization, admitted to membership or classified by any labor organization, or admitted to, or employed in, any apprenticeship or other training program, in comparison with the total number or percentage of persons of such race, color, religion, sex, or national origin in any community, State, section, or other area, or in the available work force in any community, State, section, or other area.

OTHER UNLAWFUL EMPLOYMENT PRACTICES

(a) It shall be an unlawful employment practice for an employer to discriminate against any of his employees or applicants for employment, for an employment agency, or joint labor-management committee controlling apprenticeship or other training or retraining, including on-the-job training programs, to discriminate against any individual, or for a labor organization to discriminate against any member thereof or applicant for membership, because he has opposed any practice made an unlawful employment practice by this title, or because he has made a charge, testified, assisted, or participated in any manner in an investigation, proceeding, or hearing under this title.

(b) It shall be an unlawful employment practice for an employer, labor organization, employment agency, or joint labor-management committee controlling apprenticeship or other training or retraining, including on-the-job training programs, to print or publish or cause to be printed or published any notice or advertisement relating to employment by such an employer or membership in or any classification or referral for employment by such a labor organization, or relating to any classification or referral for employment by such an employment agency, or relating to admission to, or employment in, any program established to provide apprenticeship or other training by such a joint labor-management committee, indicating any preference, limitation, specification, or discrimination, based

on race, color, religion, sex, or national origin, except that such a notice or advertisement may indicate a preference, limitation, specification, or discrimination based on religion, sex, or national origin when religion, sex, or national origin is a bona fide occupational qualification for employment.

APPENDIX F

Uniform Guidelines On Employee Selection Procedures (1978)

28 C.F.R. §50.14 (1978)
29 C.F.R. §1607 (1979)

GENERAL PRINCIPLES

1607.1. Statement of Purpose

A. Need for Uniformity—Issuing Agencies

B. Purpose of Guidelines

C. Relation to Prior Guidelines

1607.2. Scope

A. Application of Guidelines

B. Employment Decisions

C. Selection Procedures

D. Limitations

E. Indian Preference Not Affected

1607.3. Discrimination Defined: Relationship Between Use of Selection Procedures and Discrimination

A. Procedure Having Adverse Impact Constitutes Discrimination Unless Justified

B. Consideration of Suitable Alternative Selection Procedures

1607.4. Information on Impact

A. Records Concerning Impact

B. Applicable Race, Sex and Ethnic Groups For Record Keeping

C. Evaluation of Selection Rates. The "Bottom Line"

D. Adverse Impact And The "Four-Fifths Rule"

E. Consideration of User's Equal Employment Opportunity Posture

1607.5. General Standards for Validity Studies

A. Acceptable types of Validity Studies

B. Criterion-Related, Content, and Construct Validity

C. Guidelines Are Consistent with Professional Standards

D. Need For Documentation of Validity

E. Accuracy and Standardization

F. Caution Against Selection on Basis of Knowledges, Skills or Abilities Learned in Brief Orientation Period

G. Method of Use of Selection Procedures

H. Cutoff Scores

I. Use of Selection Procedures for Higher Level Jobs

J. Interim Use of Selection Procedures

K. Review of Validity Studies for Currency

1607.6. Use of Selection Procedures Which Have Not Been Validated

A. Use of Alternate Selection Procedures to Eliminate Adverse Impact

B. Where Validity Studies Cannot or Need Not Be Performed

(1) Where Informal or Unscored Procedures Are Used

(2) Where Formal And Scored Procedures Are Used

1607.7. Use of Other Validity Studies

A. Validity Studies not Conducted by the User

B. Use of Criterion-Related Validity Evidence from Other Sources

(1) Validity Evidence

(2) Job Similarity

(3) Fairness Evidence

C. Validity Evidence from Multi-Unit Study

D. Other Significant Variables

1607.8. Cooperative Studies

A. Encouragement of Cooperative Studies

B. Standards for Use of Cooperative Studies

1607.9. No Assumption of Validity

A. Unacceptable Substitutes for Evidence of Validity

B. Encouragement of Professional Supervision

1607.10. Employment Agencies and Employment Services

A. Where Selection Procedures Are Devised by Agency

B. Where Selection Procedures Are Devised Elsewhere

1607.11. Disparate Treatment

1607.12. Retesting of Applicants

1607.13. Affirmative Action

A. Affirmative Action Obligations

B. Encouragement of Voluntary Affirmative Action Programs

TECHNICAL STANDARDS

1607.14. Technical Standards for Validity Studies

A. Validity Studies Should be Based on Review of Information about the Job

B. Technical Standards for Criterion-Related Validity Studies

(1) Technical Feasibility

(2) Analysis of the Job

(3) Criterion Measures

(4) Representativeness of the Sample

(5) Statistical Relationships

(6) Operational Use of Selection Procedures

(7) Over-Statement of Validity Findings

(8) Fairness

(a) Unfairness Defined

(b) Investigation of Fairness

(c) General Considerations in Fairness Investigations

(d) When Unfairness Is Shown

(e) Technical Feasibility of Fairness Studies

(f) Continued Use of Selection Procedures When Fairness Studies not Feasible

C. Technical Standards for Content Validity Studies

(1) Appropriateness of Content Validity Studies

(2) Job Analysis for Content Validity

(3) Development of Selection Procedure

(4) Standards For Demonstrating Content Validity

(5) Reliability

(6) Prior Training or Experience

(7) Training Success

(8) Operational Use

(9) Ranking Based on Content Validity Studies

D. Technical Standards For Construct Validity Studies

(1) Appropriateness of Construct Validity Studies

(2) Job Analysis For Construct Validity Studies

(3) Relationship to the Job

(4) Use of Construct Validity Study Without New Criterion-Related Evidence

(a) Standards for Use

(b) Determination of Common Work Behaviors

DOCUMENTATION OF IMPACT AND VALIDITY EVIDENCE

1607.15. Documentation of Impact and Validity Evidence

A. Required Information

(1) Simplified Recordkeeping for Users With Less Than 100 Employees

(2) Information on Impact

(a) Collection of Information on Impact

(b) When Adverse Impact Has Been Eliminated in The Total Selection Process

(c) When Data Insufficient to Determine Impact

(3) Documentation of Validity Evidence

(a) Type of Evidence

(b) Form of Report

(c) Completeness

B. Criterion-Related Validity Studies

(1) User(s), Location(s), and Date(s) of Study

(2) Problem and Setting

(3) Job Analysis or Review of Job Information

(4) Job Titles and Codes

(5) Criterion Measures

(6) Sample Description

(7) Description of Selection Procedure

(8) Techniques and Results

(9) Alternative Procedures Investigated

(10) Uses and Applications

(11) Source Data

(12) Contact Person

(13) Accuracy and Completeness

C. Content Validity Studies

(1) User(s), Location(s), and Date(s) of Study

(2) Problem and Setting

(3) Job Analysis—Content of the Job

(4) Selection Procedure and its Content

(5) Relationship Between Selection Procedure and the Job

(6) Alternative Procedures Investigated

(7) Uses and Applications

(8) Contact Person

(9) Accuracy and Completeness

D. Construct Validity Studies

(1) User(s), Location(s), and Date(s) of Study

(2) Problem and Setting

(3) Construct Definition

(4) Job Analysis

(5) Job Titles and Codes

(6) Selection Procedure

(7) Relationship to Job Performance

(8) Alternative Procedures Investigated

(9) Uses and Applications

(10) Accuracy and Completeness

(11) Source Data

(12) Contact Person

E. Evidence of Validity from Other Studies

(1) Evidence from Criterion-Related Validity Studies

(a) Job Information

(b) Relevance of Criteria

(c) Other Variables

(d) Use of the Selection Procedure

(e) Bibliography

(2) Evidence from Content Validity Studies

(3) Evidence from Construct Validity Studies

F. Evidence of Validity from Cooperative Studies

G. Selection for Higher Level Jobs

H. Interim Use of Selection Procedures

DEFINITIONS

1607.16. Definitions

APPENDIX

1607.17. Policy Statement on Affirmative Action (see Section 13B)

1607.18. Citations

AUTHORITY: Secs. 709 and 713, Civil Rights Act of 1964 (78 Stat. 265) as amended by the Equal Employment Opportunity Act of 1972 (Pub. L. 92-261); 42 U.S.C. 2000e-8, 2000e-12.

SOURCE: 43 FR 38295 and 43 FR 38312, Aug. 25, 1978.

GENERAL PRINCIPLES

§1607.1 Statement of purpose.

A. *Need for uniformity—Issuing agencies.* The Federal government's need for a uniform set of principles on the question of the use of tests and other selection procedures has long been recognized. The Equal Employment Opportunity Commission, the Civil Service Commission, the Department of

Labor, and the Department of Justice jointly have adopted these uniform guidelines to meet that need, and to apply the same principles to the Federal Government as are applied to other employers.

B. *Purpose of guidelines.* These guidelines incorporate a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of Federal law prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin. They are designed to provide a framework for determining the proper use of tests and other selection procedures. These guidelines do not require a user to conduct validity studies of selection procedures where no adverse impact results. However, all users are encouraged to use selection procedures which are valid, especially users operating under merit principles.

C. *Relation to prior guidelines.* These guidelines are based upon and supersede previously issued guidelines on employee selection procedures. These guidelines have been built upon court decisions, the previously issued guidelines of the agencies, and the practical experience of the agencies, as well as the standards of the psychological profession. These guidelines are intended to be consistent with existing law.

§ 1607.2 Scope.

A. *Application of guidelines.* These guidelines will be applied by the Equal Employment Opportunity Commission in the enforcement of title VII of the Civil Rights Act of 1964, as amended by the Equal Employment Opportunity Act of 1972 (hereinafter "Title VII"); by the Department of Labor, and the contract compliance agencies until the transfer of authority contemplated by the President's Reorganization Plan No. 1 of 1978, in the administration and enforcement of Executive Order 11246, as amended by Executive Order 11375 (hereinafter "Executive Order 11246"); by the Civil Service Commission and

other Federal agencies subject to section 717 of Title VII; by the Civil Service Commission in exercising its responsibilities toward State and local governments under section 208(b)(1) of the Intergovernmental-Personnel Act; by the Department of Justice in exercising its responsibilities under Federal law; by the Office of Revenue Sharing of the Department of the Treasury under the State and Local Fiscal Assistance Act of 1972, as amended; and by any other Federal agency which adopts them.

B. *Employment decisions.* These guidelines apply to tests and other selection procedures which are used as a basis for any employment decision. Employment decisions include but are not limited to hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and certification, to the extent that licensing and certification may be covered by Federal equal employement opportunity law. Other selection decisions, such as selection for training or transfer, may also be considered employment decisions if they lead to any of the decisions listed above.

C. *Selection procedures.* These guidelines apply only to selection procedures which are used as a basis for making employment decisions. For example, the use of recruiting procedures designed to attract members of a particular race, sex, or ethnic group, which were previously denied employment opportunities or which are currently underutilized, may be necessary to bring an employer into compliance with Federal law, and is frequently an essential element of any effective affirmative action program; but recruitment practices are not considered by these guidelines to be selection procedures. Similarly, these guidelines do not pertain to the question of the lawfulness of a seniority system within the meaning of section 703(h), Executive Order 11246 or other provisions of Federal law or regulation, except to the extent that such systems utilize selection procedures to determine qualifications or abilities to perform the job. Nothing in these guidelines is intended or should be interpreted as discouraging the use of a selection procedure for the purpose of

determining qualifications or for the purpose of selection on the basis of relative qualifications, if the selection procedure had been validated in accord with these guidelines for each such purpose for which it is to be used.

D. *Limitations.* These guidelines apply only to persons subject to Title VII, Executive Order 11246, or other equal employment opportunity requirements of Federal law. These guidelines do not apply to responsibilities under the Age Discrimination in Employment Act of 1967, as amended, not to discriminate on the basis of age, or under sections 501, 503, and 504 of the Rehabilitation Act of 1973, not to discriminate on the basis of handicap.

E. *Indian preference not affected.* These guidelines do not restrict any obligation imposed or right granted by Federal law to users to extend a preference in employement to Indians living on or near an Indian reservation in connection with employment opportunities on or near an Indian reservation.

§ 1607.3 Discrimination defined: Relationship between use of selection procedures and discrimination.

A. *Procedure having adverse impact constitutes discrimination unless justified.* The use of any selection procedure which has an adverse impact on the hiring, promotion, or other employment or membership opportunities of members of any race, sex, or ethnic group will be considered to be discriminatory and inconsistent with these guidelines, unless the procedure has been validated in accordance with these guidelines, or the provisions of section 6 below are satisifed.

B. *Consideration of suitable alternative selection procedures.* Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact. Accordingly,

whenever a validity study is called for by these guidelines, the user should include, as a part of the validity study, an investigation of suitable alternative selection procedures and suitable alternative methods of using the selection procedure which have as little adverse impact as possible, to determine the appropriateness of using or validating them in accord with these guidelines. If a user has made a reasonable effort to become aware of such alternative procedures and validity has been demonstrated in accord with these guidelines, the use of the test or other selection procedure may continue until such time as it should reasonably be reviewed for currency. Whenever the user is shown an alternative selection procedure with evidence of less adverse impact and substantial evidence of validity for the same job in similar circumstances, the user should investigate it to determine the appropriateness of using or validating it in accord with these guidelines. This subsection is not intended to preclude the combination of procedures into a significantly more valid procedure, if the use of such a combination has been shown to be in compliance with the guidelines.

§ 1607.4 Information on impact.

A. *Records concerning impact.* Each user should maintain and have available for inspection records or other information which will disclose the impact which its tests and other selection procedures have upon employment opportunities of persons by identifiable race, sex, or ethnic group as set forth in subparagraph B below in order to determine compliance with these guidelines. Where there are large numbers of applicants and procedures are administered frequently, such information may be retained on a sample basis, provided that the sample is appropriate in terms of the applicant population and adequate in size.

B. *Applicable race, sex, and ethnic groups for recordkeeping.* The records called for by this section are to be maintained by sex, and the following races and ethnic groups: Blacks

(Negroes), American Indians (including Alaskan Natives), Asians (including Pacific Islanders), Hispanic (including persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or culture regardless of race), whites (Caucasians) other than Hispanic, and totals. The race, sex, and ethnic classifications called for by this section are consistent with the Equal Employment Opportunity Standard Form 100, Employer Information Report EEO-1 series of reports. The user should adopt safeguards to insure that the records required by this paragraph are used for appropriate purposes such as determining adverse impact, or (where required) for developing and monitoring affirmative action programs, and that such records are not used improperly. See section 4E and 17(4), below.

C. *Evaluation of selection rates. The "bottom line."* If the information called for by section 4A and B above shows that the total selection process for a job has an adverse impact, the individual components of the selection process should be evaluated for adverse impact. If this information shows that the total selection process does not have an adverse impact, the Federal enforcement agencies, in the exercise of their administrative and prosecutorial discretion, in usual circumstances, will not expect a user to evaluate the individual components for adverse impact, or to validate such individual components, and will not take enforcement action based upon adverse impact of any component of that process, including the separate parts of a multipart selection procedure or any separate procedure that is used as an alternative method of selection. However, in the following circumstances the Federal enforcement agencies will expect a user to evaluate the individual components for adverse impact and may, where appropriate, taken enforcement action with respect to the individual components for adverse impact and may, where appropriate, taken enforcement action with respect to the individual components: (1) where the selection procedure is a significant factor in the continuation

of patterns of assignments of incumbent employees caused by prior discriminatory employment practices, (2) where the weight of court decisions or administrative interpretations hold that a specific procedure (such as height or weight requirements or no-arrest records) is not job related in the same or similar circumstances. In unusual circumstances, other than those listed in (1) and (2) above, the Federal enforcement agencies may request a user to evaluate the individual components for adverse impact and may, where appropriate, take enforcement action with respect to the individual component.

D. *Adverse impact and the "four-fifths rule."* A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group. Where the user's evidence concerning the impact of a selection procedure indicates adverse impact but is based upon numbers which are too small to be reliable, evidence concerning the impact of the procedure over a longer period of time and/or evidence concerning the impact which the selection procedure had when used in the same manner in similar circumstances elsewhere may be considered in determining adverse impact. Where the user has not maintained data on adverse impact as required by the documentation section of applicable guidelines, the Federal enforcement

agencies may draw an inference of adverse impact of the selection process from the failure of the user to maintain such data, if the user has an underutilization of a group in the job category, as compared to the group's representation in the relevant labor market or, in the case of jobs filled from within, the applicable work force.

E. *Consideration of user's equal employment opportunity posture.* In carrying out their obligations, the Federal enforcement agencies will consider the general posture of the user with respect to equal employment opportunity for the job or group of jobs in question. Where a user has adopted an affirmative action program, the Federal enforcement agencies will consider the provisions of that program, including the goals and timetables which the user has adopted and the progress which the user has made in carrying out that program and in meeting the goals and timetables. While such affirmative action programs may in design and execution be race, color, sex, or ethnic conscious, selection procedures under such programs should be based upon the ability or relative ability to do the work.

§ 1607.5 General standards for validity studies.

A. *Acceptable types of validity studies.* For the purposes of satisfying these guidelines, users may rely upon criterion-related validity studies, content validity studies or construct validity studies, in accordance with the standards set forth in the technical standards of these guidelines, section 14 below. New strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession.

B. *Criterion-related, content, and construct validity.* Evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job per-

formance. See section 14B below. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. See 14C below. Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated. See section 14D below.

C. *Guidelines are consistent with professional standards.* The provisions of these guidelines relating to validation of selection procedures are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures, such as those described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (American Psychological Association, Washington, D.C., 1974) (hereinafter "A.P.A. Standards") and standard textbooks and journals in the field of personnel selection.

D. *Need for documentation of validity.* For any selection procedure which is part of a selection process which has an adverse impact and which selection procedure has an adverse impact, each user should maintain and have available such documentation as is described in section 15 below.

E. *Accuracy and standardization.* Validity studies should be carried out under conditions which assure insofar as possible the adequacy and accuracy of the research and the report. Selection procedures should be administered and scored under standardized conditions.

F. *Caution against selection on basis of knowledges, skills, or ability learned in brief orientation period.* In general, users should avoid making employment decisions on the basis of measures of knowledges, skills, or abilities which are normally learned in a brief orientation period, and which have an adverse impact.

G. *Method of use of selection procedures.* The evidence of both the validity and utility of a selection procedure should support the method the user chooses for operational use of the procedure, if that method of use has a greater adverse impact than another method of use. Evidence which may be sufficient to support the use of a selection procedure on a pass/fail (screening) basis may be insufficient to support the use of the same procedure on a ranking basis under these guidelines. Thus, if a user decides to use a selection procedure on a ranking basis, and that method of use has a greater adverse impact than use on an appropriate pass/fail basis (see section 5H below), the user should have sufficient evidence of validity and utility to support the use on a ranking basis. See sections 3B, 14B (5) and (6), and 14C (8) and (9).

H. *Cutoff scores.* Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force. Where applicants are ranked on the basis of properly validated selection procedures and those applicants scoring below a higher cutoff score than appropriate in light of such expectations have little or no chance of being selected for employment, the higher cutoff score may be appropriate, but the degree of adverse impace should be considered.

I. *Use of selection procedures for higher level jobs.* If job progression structures are so established that employees will probably, within a reasonable period of time and in a majority of cases, progress to a higher level, it may be considered that the applicants are being evaluated for a job or jobs at the higher

level. However, where job progression is not so nearly automatic, or the time span is such that higher level jobs or employees' potential may be expected to change in significant ways, it should be considered that applicants are being evaluated for a job at or near the entry level. A "reasonable period of time" will vary for different jobs and employment situations but will seldom be more than 5 years. Use of selection procedures to evaluate applicants for a higher level job would not be appropriate:

(1) If the majority of those remaining employed do not progress to the higher level job;

(2) If there is a reason to doubt that the higher level job will continue to require essentially similar skills during the progression period; or

(3) If the selection procedures measure knowledges, skills, or abilities required for advancement which would be expected to develop principally from the training or experience on the job.

J. *Interim use of selection procedures.* Users may continue the use of a selection procedure which is not at the moment fully supported by the required evidence of validity, provided: (1) The user has available substantial evidence of validity, and (2) the user has in progress, when technically feasible, a study which is designed to produce the additional evidence required by these guidelines within a reasonable time. If such a study is not technically feasible, see section 6B. If the study does not demonstrate validity, this provision of these guidelines for interim use shall not constitute a defense in any action, nor shall it relieve the user of any obligations arising under Federal law.

K. *Review of validity studies for currency.* Whenever validity has been shown in accord with these guidelines for the use of a particular selection procedure for a job or group of jobs, additional studies need not be performed until such time as the validity study is subject to review as provided in section 3B

above. There are no absolutes in the area of determining the currency of a validity study. All circumstances concerning the study, including the validation strategy used, and changes in the relevant labor market and the job should be considered in the determination of when a validity study is outdated.

§ 1607.6 Use of selection procedures which have not been validated.

A. *Use of alternate selection procedures to eliminate adverse impact.* A user may choose to utilize alternative selection procedures in order to eliminate adverse impact or as part of an affirmative action program. See section 13 below. Such alternative procedures should eliminate the adverse impact in the total selection process, should be lawful and should be as job related as possible.

B. *Where validity studies cannot or need not be performed.* There are circumstances in which a user cannot or need not utilize the validation techniques contemplated by these guidelines. In such circumstances, the user should utilize selection procedures which are as job related as possible and which will minimize or eliminate adverse impact, as set forth below.

(1) *Where informal or unscored procedures are used.* When an informal or unscored selection procedure which has an adverse impact is utilized, the user should eliminate the adverse impact, or modify the procedure to one which is a formal, scored or quantified measure or combination of measures and then validate the procedure in accord with these guidelines, or otherwise justify continued use of the procedure in accord with Federal law.

(2) *Where formal and scored procedures are used.* When a formal and scored selection procedure is used which has an adverse impact, the validation techniques contemplated by these guidelines usually should be followed if technically feasible. Where the user cannot or need not follow the validation techni-

ques anticipated by these guidelines, the user should either modify the procedure to eliminate adverse impact or otherwise justify continued use of the procedure in accord with Federal law.

§ 1607.7 Use of other validity studies.

A. *Validity studies not conducted by the user.* Users may, under certain circumstances, support the use of selection procedures by validity studies conducted by other users or conducted by test publishers or distributors and described in test manuals. While publishers of selection procedures have a professional obligation to provide evidence of validity which meets generally accepted professional standards (see section 5C above), users are cautioned that they are responsible for compliance with these guidelines. Accordingly, users seeking to obtain selection procedures from publishers and distributors should be careful to determine that, in the event the user becomes subject to the validity requirements of these guidelines, the necessary information to support validity has been determined and will be made available to the user.

B. *Use of criterion-related validity evidence from other sources.* Criterion-related validity studies conducted by one test user, or described in test manuals and the professional literature, will be considered acceptable for use by another user when the following requirements are met:

(1) *Validity evidence.* Evidence from the available studies meeting the standards of section 14B below clearly demonstrates that the selection procedure is valid;

(2) *Job similarity.* The incumbents in the user's job and the incumbents in the job or group of jobs on which the validity study was conducted perform substantially the same major work behaviors, as shown by appropriate job analyses both on the job or group of jobs on which the validity study was per-

formed and on the job for which the selection procedure is to be used; and

(3) *Fairness evidence.* The studies include a study of test fairness for each race, sex, and ethnic group which constitutes a significant factor in the borrowing user's relevant labor market for the job or jobs in question. If the studies under consideration satisfy (1) and (2) above but do not contain an investigation of test fairness, and it is not technically feasible for the borrowing user to conduct an internal study of test fairness, the borrowing user may utilize the study until studies conducted elsewhere meeting the requirements of these guidelines show test unfairness, or until such time as it becomes technically feasible to conduct an internal study of test fairness and the results of that study can be acted upon. Users obtaining selection procedures from publishers should consider, as one factor in the decision to purchase a particular selection procedure, the availability of evidence concerning test fairness.

C. *Validity evidence from multiunit study.* If validity evidence from a study covering more than one unit within an organization statisfies the requirements of section 14B below, evidence of validity specific to each unit will not be required unless there are variables which are likely to affect validity significantly.

D. *Other significant variables.* If there are variables in the other studies which are likely to affect validity significantly, the user may not rely upon such studies, but will be expected either to conduct an internal validity study or to comply with section 6 above.

§ 1607.8 Cooperative studies.

A. *Encouragement of cooperative studies.* The agencies issuing these guidelines encourage employers, labor organizations, and employment agencies to cooperate in research, development, search for lawful alternatives, and validity studies in

order to achieve procedures which are consistent with these guidelines.

B. *Standards for use of cooperative studies.* If validity evidence from a cooperative study satisfies the requirements of section 14 below, evidence of validity specific to each user will not be required unless there are variables in the user's situation which are likely to affect validity significantly.

§ 1607.9 No assumption of validity.

A. *Unacceptable substitutes for evidence of validity.* Under no circumstances will the general reputation of a test or other selection procedures, its author or its publisher, or casual reports of it's validity be accepted in lieu of evidence of validity. Specifically ruled out are: assumptions of validity based on a procedure's name or descriptive labels; all forms of promotional literature; data bearing on the frequency of a procedure's usage; testimonial statements and credentials of sellers, users, or consultants; and other nonempirical or anecdotal accounts of selection practices or selection outcomes.

B. *Encouragement of professional supervision.* Professional supervision of selection activities is encouraged but is not a substitute for documented evidence of validity. The enforcement agencies will take into account the fact that a thorough job analysis was conducted and that careful development and use of a selection procedure in accordance with professional standards enhance the probability that the selection procedure is valid for the job.

§ 1607.10 Employment agencies and employment services.

A. *Where selection procedures are devised by agency.* An employment agency, including private employment agencies and State employment agencies, which agrees to a request by an employer or labor organization to device and utilize a selection procedure should follow the standards in these guidelines for

determining adverse impact. If adverse impact exists the agency should comply with these guidelines. An employment agency is not relieved of its obligation herein because the user did not request such validation or has requested the use of some lesser standard of validation than is provided in these guidelines. The use of an employment agency does not relieve an employer or labor organization or other user of its responsibilities under Federal law to provide equal employment opportunity or its obligations as a user under these guidelines.

B. *Where selection procedures are devised elsewhere.* Where an employment agency or service is requested to administer a selection procedure which has been devised elsewhere and to make referrals pursuant to the results, the employment agency or service should maintain and have available evidence of the impact of the selection and referral procedures which it administers. If adverse impact results the agency or service should comply with these guidelines. If the agency or service seeks to comply with these guidelines by reliance upon validity studies or other data in the possession of the employer, it should obtain and have available such information.

§ 1607.11 Disparate treatment.

The principles of disparate or unequal treatment must be distinguished from the concepts of validation. A selection procedure—even though validated against job performance in accordance with these guidelines—cannot be imposed upon members of a race, sex, or ethnic group where other employees, applicants, or members have not been subject to that standard. Disparate treatment occurs where members of a race, sex, or ethnic group have been denied the same employment, promotion, membership, or other employment opportunities as have been available to other employees or applicants. Those employees or applicants who have been denied equal treatment, because of prior discriminatory practices or policies, must at least be afforded the same opportunities as had existed for other

employees or applicants during the period of discrimination. Thus, the persons who were in the class of persons discriminated against during the period the user followed the discriminatory practices should be allowed the opportunity to qualify under less stringent selection procedures previously followed, unless the user demonstrates that the increased standards are required by business necessity. This section does not prohibit a user who has not previously followed merit standards from adopting merit standards which are in compliance with these guidelines; nor does it preclude a user who has previously used invalid or unvalidated selection procedures from developing and using procedures which are in accord with these guidelines.

§ 1607.12 Retesting of applicants.

Users should provide a reasonable opportunity for retesting and reconsideration. Where examinations are administered periodically with public notice, such reasonable opportunity exists, unless persons who have previously been tested are precluded from retesting. The user may however take reasonable steps to preserve the security of its procedures.

§ 1607.13 Affirmative action.

A. Affirmative action obligations. The use of selection procedures which have been validated pursuant to these guidelines does not relieve users of any obligations they may have to undertake affirmative action to assure equal employment opportunity. Nothing in these guidelines is intended to preclude the use of lawful selection procedures which assist in remedying the effects of prior discriminatory practices, or the achievement of affirmative action objectives.

B. *Encouragement of voluntary affirmative action programs.* These guidelines are also intended to encourage the adoption and implementation of voluntary affirmative action programs by users who have no obligation under Federal law to adopt them; but are not intended to impose any new obligations in

that regard. The agencies issuing and endorsing these guidelines endorse for all private employers and reaffirm for all governmental employers the Equal Employment Opportunity Coordinating Council's "Policy Statement on Affirmative Action Programs for State and Local Government Agencies" (41 FR 38814, September 13, 1976). That policy statement is attached hereto as appendix, section 17.

TECHNICAL STANDARDS

§1607.14 Technical standards for validity studies.

The following minimum standards, as applicable, should be met in conducting a validity study. Nothing in these guidelines is intended to preclude the development and use of other professionally acceptable techniques with respect to validation of selection procedures. Where it is not technically feasible for a user to conduct a validity study, the user has the obligation otherwise to comply with these guidelines. See sections 6 and 7 above.

A. *Validity studies should be based on review of information about the job.* Any validity study should be based upon a review of information about the job for which the selection procedure is to be used. The review should include a job analysis except as provided in section 14B(3) below with respect to criterion-related validity. Any method of job analysis may be used if it provides the information required for the specific validation strategy used.

B. *Technical standards for criterion-related validity studies.—(1) Technical feasibility.* Users choosing to validate a selection procedure by a criterion-related validity strategy should determine whether it is technically feasible (as defined in section 16) to conduct such a study in the particular employment context. The determination of the number of persons necessary to permit the conduct of a meaningful criterion-related study should be made by the user on the basis of all rele-

vant information concerning the selection procedure, the potential sample and employment situation. Where appropriate, jobs with substantially the same major work behaviors may be grouped together for validity studies, in order to obtain an adequate sample. These guidelines do not require a user to hire or promote persons for the purpose of making it possible to conduct a criterion-related study.

(2) *Analysis of the job.* There should be a review of job information to determine measures of work behavior(s) or performance that are relevant to the job or group of jobs in question. These measures or criteria are relevant to the extent that they represent critical or important job duties, work behaviors or work outcomes as developed from the review of job information. The possibility of bias should be considered both in selection of the criterion measures and their application. In view of the possibility of bias in subjective evaluations, supervisory rating techniques and instructions to raters should be carefully developed. All criterion measures and the methods for gathering data need to be examined for freedom from factors which would unfairly alter scores of members of any group. The relevance of criteria and their freedom from bias are of particular concern when there are significant differences in measures of job performance for different groups.

(3) *Criterion measures.* Proper safeguards should be taken to insure that scores on selection procedures do not enter into any judgments of employee adequacy that are to be used as criterion measures. Whatever criteria are used should represent important or critical work behavior(s) or work out-comes. Certain criteria may be used without a full job analysis if the user can show the importance of the criteria to the particular employment context. These criteria include but are not limited to production rate, error rate, tardiness, absenteeism, and length of service. A standardized rating of overall work performance may be used where a study of the job shows that it is an appropriate criterion. Where performance in training is used as a criterion,

success in training should be properly measured and the relevance of the training should be shown either through a comparison of the content of the training program with the critical or important work behavior(s) of the job(s), or through a demonstration of the relationship between measures of performance in training and measures of job performance. Measures of relative success in training include but are not limited to instructor evaluations, performance samples, or tests. Criterion measures consisting of paper and pencil tests will be closely reviewed for job relevance.

(4) *Representativeness of the sample.* Whether the study is predictive or concurrent, the sample subjects should insofar as feasible be representative of the candidates normally available in the relevant labor market for the job or group of jobs in question, and should insofar as feasible include the races, sexes, and ehtnic groups normally available in the relevant job market. In determining the representativeness of the sample in a concurrent validity study, the user should take into account the extent to which the specific knowledges or skills which are the primary focus of the test are those which employees learn on the job.

Where samples are combined or compared, attention should be given to see that such samples are comparable in terms of the actual job they perform, the length of time on the job where time on the job is likely to affect performance, and other relevant factors likely to affect validity differences; or that these factors are included in the design of the study and their effects identified.

(5) *Statistical relationships.* The degree of relationship between selection procedure scores and criterion measures should be examined and computed, using professionally acceptable statistical procedures. Generally, a selection procedure is considered related to the criterion, for the purposes of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is

statistically significant at the 0.05 level of significance, which means that it is sufficiently high as to have a probability of no more than one (1) in twenty (20) to have occurred by chance. Absence of a statistically significant relationship between a selection procedure and job performance should not necessarily discourage other investigations of the validity of that selection procedure.

(6) *Operational use of selection procedures.* Users should evaluate each selection procedure to assure that it is appropriate for operational use, including establishment of cutoff scores or rank ordering. Generally, if other factors remain the same, the greater the magnitude of the relationship (e.g., correlation coefficient) between performance on a selection procedure and one or more criteria of performance on the job, and the greater the importance and number of aspects of job performance covered by the criteria, the more likely it is that the procedure will be appropriate for use. Reliance upon a selection procedure which is significantly related to a criterion measure, but which is based upon a study involving a large number of subjects and has a low correlation coefficient will be subject to close review if it has a large adverse impact. Sole reliance upon a single selection instrument which is related to only one of many job duties or aspects of job performance will also be subject to close review. The appropriateness of a selection procedure is best evaluated in each particular situation and there are no minimum correlation coefficients applicable to all employment situations. In determining whether a selection procedure is appropriate for operational use the following considerations should also be taken into account: The degree of adverse impact of the procedure, the availability of other selection procedures of greater or substantially equal validity.

(7) *Overstatement of validity findings.* Users should avoid reliance upon techniques which tend to overestimate validity findings as a result of capitalization on chance unless an appropriate safeguard is taken. Reliance upon a few selection pro-

cedures or criteria of successful job performance when many selection procedures or criteria of performance have been studied, or the use of optimal statistical weights for selection procedures computed in one sample, are techniques which tend to inflate validity estimates as a result of chance. Use of a large sample is one safeguard: cross-validation is another.

(8) *Fairness.* This section generally calls for studies of unfairness where technically feasible. The concept of fairness or unfairness of selection procedures is a developing concept. In addition, fairness studies generally require substantial numbers of employees in the job or group of jobs being studied. For these reasons, the Federal enforcement agencies recognize that the obligation to conduct studies of fairness imposed by the guidelines generally will be upon users or groups of users with a large number of persons in a job class, or test developers; and that small users utilizing their own selection procedures will generally not be obligated to conduct such studies because it will be technically infeasible for them to do so.

(a) *Unfairness defined.* When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores.

(b) *Investigation of fairness.* Where a selection procedure results in an adverse impact on a race, sex, or ethnic group identified in accordance with the classifications set forth in section 4 above and that group is a significant factor in the relevant labor market, the user generally should investigate the possible existence of unfairness for that group if it is technically feasible to do so. The greater the severity of the adverse impact on a group, the greater the need to investigate the possible existence of unfairness. Where the weight of evidence from other studies shows

that the selection procedure predicts fairly for the group in question and for the same or similar jobs, such evidence may be relied on in connection with the selection procedure at issue.

(c) *General considerations in fairness investigations.* Users conducting a study of fairness should review the A.P.A. Standards regarding investigation of possible bias in testing. An investigation of fairness of a selection procedure depends on both evidence of validity and the manner in which the selection procedure is to be used in a particular employment context. Fairness of a selection procedure cannot necessarily be specified in advance without investigating these factors. Investigation of fairness of a selection procedure in samples where the range of scores on selection procedures or criterion measures is severely restricted for any subgroup sample (as compared to other subgroup samples) may produce misleading evidence of unfairness. That factor should accordingly be taken into account in conducting such studies and before reliance is placed on the results.

(d) *When unfairness is shown.* If unfairness is demonstrated through a showing that members of a particular group perform better or poorer on the job than their scores on the selection procedure would indicate through comparison with how members of other groups perform, the user may either revise or replace the selection instrument in accordance with these gı idelines, or may continue to use the selection instrument operationally with appropriate revisions in its use to assure compatibility between the probability of successful job performance and the probability of being selected.

(e) *Technical feasibility of fairness studies.* In addition to the general conditions needed for technical feasibility for the conduct of a criterion-related study (see section 16, below) an investigation of fairness requires the following:

(i) An adequate sample of persons in each group available for the study to achieve findings of statistical significance.

Guidelines do not require a user to hire or promote persons on the basis of group classifications for the purpose of making it possible to conduct a study of fairness; but the user has the obligation otherwise to comply with these guidelines.

(ii) The samples for each group should be comparable in terms of the actual job they perform, length of time on the job where time on the job is likely to affect performance, and other relevant factors likely to affect validity differences; or such factors should be included in the design of the study and their effects identified.

(f) *Continued use of selection procedures when fairness studies not feasible.* If a study of fairness should otherwise be performed, but is not technically feasible, a selection procedure may be used which has otherwise met the validity standards of these guidelines, unless the technical infeasibility resulted from discriminatory employment practices which are demonstrated by facts other than past failure to conform with requirements for validation of selection procedures. However, when it becomes technically feasible for the user to perform a study of fairness and such a study is otherwise called for, the user should conduct the study of fairness.

C. *Technical standards for content validity studies.*—(1) *Appropriateness of content validity studies.* Users choosing to validate a selection procedure by a content validity strategy should determine whether it is appropriate to conduct such a study in the particular employment context. A selection procedure can be supported by a content validity strategy to the extent that it is a representative sample of the content of the job. Selection procedures which purport to measure knowledges, skills, or abilities may in certain circumstances be justified by content validity, although they may not be representative samples, if the knowledge, skill, or ability measured by the selection procedure can be operationally defined as provided in section 14C(4) below, and if that knowledge, skill, or ability is a necessary prerequisite to successful job performance.

A selection procedure based upon inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs, such as intelligence, aptitude, personality, commonsense, judgment, leadership, and spatial ability. Content validity is also not an appropriate strategy when the selection procedure involves knowledges, skills, or abilities which an employee will be expected to learn on the job.

(2) *Job analysis for content validity.* There should be a job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance and, if the behavior results in work product(s), an analysis of the work product(s). Any job analysis should focus on the work behavior(s) and the tasks associated with them. If work behavior(s) are not observable, the job analysis should identify and analyze those aspects of the behavior(s) that can be observed and the observed work products. The work behavior(s) selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.

(3) *Development of selection procedures.* A selection procedure designed to measure the work behavior may be developed specifically from the job and job analysis in question, or may have been previously developed by the user, or by other users or by a test publisher.

(4) *Standards for demonstrating content validity.* To demonstrate the content validity of a selection procedure, a user should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job. In the case of a selection procedure measuring a knowledge, skill or

ability, the knowledge, skill, or ability being measured should be operationally defined. In the case of a selection procedure measuring a knowledge, the knowledge being measured should be operationally defined as that body of learned information which is used in and is a necessary prerequisite for observable aspects of work behavior of the job. In the case of skills or abilities, the skill or ability being measured should be operationally defined in terms of observable aspects of work behavior of the job. For any selection procedure measuring a knowledge, skill, or ability the user should show that (a) the selection procedure measures and is a representative sample of that knowledge, skill, or ability; and (b) that knowledge, skill, or ability is used in and is a necessary prerequisite to performance of critical or important work behavior(s). In addition, to be content valid, a selection procedure measuring a skill or ability should either closely approximate an observable work behavior, or its product should closely approximate an observable work product. If a test purports to sample a work behavior or to provide a sample of a work product, the manner and setting of the selection procedure and its level and complexity should closely approximate the work situation. The closer the content and the context of the selection procedure are to work samples or work behaviors, the stronger is the basis for showing content validity. As the content of the selection procedure less resembles a work behavior, or the setting and manner of the administration of the selection procedure less resemble the work situation, or the result less resembles a work product, the less likely the selection procedure is to be content valid, and the greater the need for other evidence of validity.

(5) *Reliability.* The reliability of selection procedures justified on the basis of content validity should be a matter of concern to the user. Whenever it is feasible, appropriate statistical estimates should be made of the reliability of the selection procedure.

(6) *Prior training or experience.* A requirement for or evaluation of specific prior training or experience based on content validity, including a specification of level or amount of training or experience, should be justified on the basis of the relationship between the content of the training or experience and the content of the job for which the training or experience is to be required or evaluated. The critical consideration is the resemblance between the specific behaviors, products, knowledges, skills, or abilities in the experience or training and the specific behaviors, products, knowledges, skills, or abilities required on the job, whether or not there is close resemblance between the experience or training as a whole and the job as a whole.

(7) *Content validity of training success.* Where a measure of success in a training program is used as a selection procedure and the content of a training program is justified on the basis of content validity, the use should be justified on the relationship between the content of the training program and the content of the job.

(8) *Operational use.* A selection procedure which is supported on the basis of content validity may be used for a job if it represents a critical work behavior (i.e., a behavior which is necessary for performance of the job) or work behaviors which constitute most of the important parts of the job.

(9) *Ranking based on content validity studies.* If a user can show, by a job analysis or otherwise, that a higher score on a content valid selection procedure is likely to result in better job performance, the results may be used to rank persons who score above minimum levels. Where a selection procedure supported solely or primarily by content validity is used to rank job candidates, the selection procedure should measure those aspects of performance which differentiate among levels of job performance.

D. *Technical standards for construct validity studies.*—(1) *Appropriateness of construct validity studies.* Construct validity is a more complex strategy than either criterion-related or content validity. Construct validation is a relatively new and developing procedure in the employment field, and there is at present a lack of substantial literature extending the concept to employment practices. The user should be aware that the effort to obtain sufficient empirical support for construct validity is both an extensive and arduous effort involving a series of research studies, which include criterion related validity studies and which may include content validity studies. Users choosing to justify use of a selection procedure by this strategy should therefore take particular care to assure that the validity study meets the standards set forth below.

(2) *Job analysis for construct validity studies.* There should be a job analysis. This job analysis should show the work behavior(s) required for successful performance of the job, or the groups of jobs being studied, the critical or important work behavior(s) in the job or group of jobs being studied, and an identification of the construct(s) believed to underlie successful performance of these critical or important work behaviors in the job or jobs in question. Each construct should be named and defined, so as to distinguish it from other constructs. If a group of jobs is being studied the jobs should have in common one or more critical or important work behaviors at a comparable level of complexity.

(3) *Relationship to the job.* A selection procedure should then be identified or developed which measures the construct identified in accord with subparagraph (2) above. The user should show by empirical evidence that the selection procedure is validly related to the construct and that the construct is validly related to the performance of critical or important work behavior(s). The relationship between the construct as measured by the selection procedure and the related work behavior(s) should be supported by empirical evidence from one or more

criterion-related studies involving the job or jobs in question which satisfy the provisions of section 14B above.

(4) *Use of construct validity study without new criterion-related evidence.*—(a) *Standards for use.* Until such time as professional literature provides more guidance on the use of construct validity in employment situations, the Federal agencies will accept a claim of construct validity without a criterion-related study which satisfies section 14B above only when the selection procedure has been used elsewhere in a situation in which a criterion-related study has been conducted and the use of a criterion-related validity study in this context meets the standards for transportability of criterion-related validity studies as set forth above in section 7. However, if a study pertains to a number of jobs having common critical or important work behaviors at a comparable level of complexity, and the evidence satisfies subparagraphs 14B (2) and (3) above for those jobs with criterion-related validity evidence for those jobs, the selection procedure may be used for all the jobs to which the study pertains. If construct validity is to be generalized to other jobs or groups of jobs not in the group studied, the Federal enforcement agencies will expect at a minimum additional empirical research evidence meeting the standards of subparagraphs section 14B (2) and (3) above for the additional jobs or groups of jobs.

(b) *Determination of common work behaviors.* In determining whether two or more jobs have one or more work behavior(s) in common, the user should compare the observed work behavior(s) in each of the jobs and should compare the observed work product(s) in each of the jobs. If neither the observed work behavior(s) in each of the jobs nor the observed work product(s) in each of the jobs are the same, the Federal enforcement agencies will presume that the work behavior(s) in each job are different. If the work behaviors are not observable, then evidence of similarity of work products and any other rele-

vant research evidence will be considered in determining whether the work behavior(s) in the two jobs are the same.

DOCUMENTATION OF IMPACT AND VALIDITY EVIDENCE

§1607.15 Documentation of impact and validity evidence.

A. *Required information.* Users of selection procedures other than those users complying with section 15A(1) below should maintain and have available for each job information on adverse impact of the selection process for that job and, where it is determined a selection process has an adverse impact, evidence of validity as set forth below.

(1) *Simplified recordkeeping for users with less than 100 employees.* In order to minimize recordkeeping burdens on employers who employ one hundred (100) or fewer employees, and other users not required to file EEO-1, et seq., reports, such users may satisfy the requirements of this section 15 if they maintain and have available records showing, for each year:

(a) The number of persons hired, promoted, and terminated for each job, by sex, and where appropriate by race and national origin;

(b) The number of applicants for hire and promotion by sex and where appropriate by race and national origin; and

(c) The selection procedures utilized (either standardized or not standardized).

These records should be maintained for each race or national origin group (see section 4 above) constituting more than two percent (2%) of the labor force in the relevant labor area. However, it is not necessary to maintain records by race and/or national origin (see § 4 above) if one race or national origin group in the relevant labor area constitutes more than ninety-eight percent (98%) of the labor force in the area. If the user has

reason to believe that a selection procedure has an adverse impact, the user should maintain any available evidence of validity for that procedure (see sections 7A and 8).

(2) *Information on impact.*—(a) *Collection of information on impact.* Users of selection procedures other than those complying with section 15A(1) above should maintain and have available for each job records or other information showing whether the total selection process for that job has an adverse impact on any of the groups for which records are called for by sections 4B above. Adverse impact determinations should be made at least annually for each such group which constitutes at least 2 percent of the labor force in the relevant labor area or 2 percent of the applicable workforce. Where a total selection process for a job has an adverse impact, the user should maintain and have available records or other information showing which components have an adverse impact. Where the total selection process for a job does not have an adverse impact, information need not be maintained for individual components except in circumstances set forth in subsection 15A(2)(b) below. If the determination of adverse impact is made using a procedure other than the "four-fifths rule," as defined in the first sentence of section 4D above, a justification, consistent with section 4D above, for the procedure used to determine adverse impact should be available.

(b) *When adverse impact has been eliminated in the total selection process.* Whenever the total selection process for a particular job has had an adverse impact, as defined in section 4 above, in any year, but no longer has an adverse impact, the user should maintain and have available the information on individual components of the selection process required in the preceding paragraph for the period in which there was adverse impact. In addition, the user should continue to collect such information for at least two (2) years after the adverse impact has been eliminated.

(c) *When data insufficient to determine impact.* Where there has been an insufficient number of selections to determine whether there is an adverse impact of the total selection process for a particular job, the user should continue to collect, maintain and have available the information on individual components of the selection process required in section 15(A)(2)(a) above until the information is sufficient to determine that the overall selection process does not have an adverse impact as defined in section 4 above, or until the job has changed substantially.

(3) *Documentation of validity evidence.*—(a) *Types of evidence.* Where a total selection process has an adverse impact (see section 4 above) the user should maintain and have available for each component of that process which has an adverse impact, one or more of the following types of documentation evidence:

(i) Documentation evidence showing criterion-related validity of the selection procedure (see section 15B, below).

(ii) Documentation evidence showing content validity of the selection procedure (see section 15C, below).

(iii) Documentation evidence showing construct validity of the selection procedure (see section 15D, below).

(iv) Documentation evidence from other studies showing validity of the selection procedure in the user's facility (see section 15E, below).

(v) Documentation evidence showing why a validity study cannot or need not be performed and why continued use of the procedure is consistent with Federal law.

(b) *Form of report.* This evidence should be compiled in a reasonably complete and organized manner to permit direct evaluation of the validity of the selection procedure. Previously written employer or consultant reports of validity, or reports

describing validity studies completed before the issuance of these guidelines are acceptable if they are complete in regard to the documentation requirements contained in this section, or if they satisfied requirements of guidelines which were in effect when the validity study was completed. If they are not complete, the required additional documentation should be appended. If necessary information is not available the report of the validity study may still be used as documentation, but its adequacy will be evaluated in terms of compliance with the requirements of these guidelines.

(c) *Completeness.* In the event that evidence of validity is reviewed by an enforcement agency, the validation reports completed after the effective date of these guidelines are expected to contain the information set forth below. Evidence denoted by use of the word "(Essential)" is considered critical. If information denoted essential is not included, the report will be considered incomplete unless the user affirmatively demonstrates either its unavailability due to circumstances beyond the user's control or special circumstances of the user's study which make the information irrelevant. Evidence not so denoted is desirable but its absence will not be a basis for considering a report incomplete. The user should maintain and have available the information called for under the heading "Source Data" in sections 15B(11) and 15D(11). While it is a necessary part of the study, it need not be submitted with the report. All statistical results should be organized and presented in tabular or graphic form to the extent feasible.

B. *Criterion-related validity studies.* Reports of criterion-related validity for a selection procedure should include the following information:

(1) *User(s), location(s), and date(s) of study.* Dates and location(s) of the job analysis or review of job information, the date(s) and location(s) of the administration of the selection procedures and collection of criterion data, and the time between collection of data on selection procedures and criterion

measures should be provided (Essential). If the study was conducted at several locations, the address of each location, including city and State, should be shown.

(2) *Problem and setting.* An explicit definition of the purpose(s) of the study and the circumstances in which the study was conducted should be provided. A description of existing selection procedures and cutoff scores, if any, should be provided.

(3) *Job anlysis or review of job information.* A description of the procedure used to analyze the job or group of jobs, or to review the job information should be provided (Essential). Where a review of job information results in criteria which may be used without a full job analysis (see section 14B(3)), the basis for the selection of these criteria should be reported (Essential). Where a job analysis is required a complete description of the work behavior(s) or work outcome(s), and measures of their criticality or importance should be provided (Essential). The report should describe the basis on which the behavior(s) or outcome(s) were determined to be critical or important, such as the proportion of time spent on the respective behaviors, their level of difficulty, their frequencey of performance, the consequences of error, or other appropriate factors (Essential). Where two or more jobs are grouped for a validity study, the information called for in this subsection should be provided for each of the jobs, and the justification for the grouping (see section 14B(1)) should be provided (Essential).

(4) *Job titles and codes.* It is desirable to provide the user's job title(s) for the job(s) in question and the corresponding job title(s) and code(s) from U.S. Employment Service's Dictionary of Occupational Titles.

(5) *Criterion measures.* The bases for the selection of the criterion measures should be provided, together with references to the evidence considered in making the selection of criterion measures (essential). A full description of all criteria on which

data were collected and means by which they were observed, recorded, evaluated, and quantified, should be provided (essential). If rating techniques are used as criterion measures, the appraisal form(s) and instructions to the rater(s) should be included as part of the validation evidence, or should be explicitly described and available (essential). All steps taken to insure that criterion measures are free from factors which would unfairly alter the scores of members of any group should be described (essential).

(6) *Sample description.* A description of how the research sample was identified and selected should be included (essential). The race, sex, and ethnic composition of the sample, including those groups set forth in section 4A above, should be described (essential). This description should include the size of each subgroup (essential). A description of how the research sample compares with the relevant labor market or work force, the method by which the relevant labor market or work force was defined, and a discussion of the likely effects on validity of differences between the sample and the relevant labor market or work force, are also desirable. Descriptions of educational levels, length of service, and age are also desirable.

(7) *Description of selection procedures.* Any measure, combination of measures, or procedure studied should be completely and explicitly described or attached (essential). If commercially available selection procedures are studied, they should be described by title, form, and publisher (essential). Reports of reliability estimates and how they were established are desirable.

(8) *Techniques and results.* Methods used in analyzing data should be described (essential). Measures of central tendency (e.g., means) and measures of dispersion (e.g., standard deviations and ranges) for all selection procedures and all criteria should be reported for each race, sex, and ethnic group which constitutes a significant factor in the relevant labor market (essential). The magnitude and direction of all relationships between selection procedures and criterion measures investigated

should be reported for each relevant race, sex, and ethnic group and for the total group (essential). Where groups are too small to obtain reliable evidence of the magnitude of the relationship, need not be reported separately. Statements regarding the statistical significance of results should be made (essential). Any statistical adjustments, such as for less then perfect reliability or for restriction of score range in the selection procedure or criterion should be described and explained; and uncorrected correlation coeffients should also be shown (essential). Where the statistical technique categorizes continuous data, such as biserial correlation and the phi coefficient, the categories and the bases on which they were determined should be described and explained (essential). Studies of test fairness should be included where called for by the requirements of section 14B(8) (essential). These studies should include the rationale by which a selection procedure was determined to be fair to the group(s) in question. Where test fairness or unfairness has been demonstrated on the basis of other studies, a bibliography of the relevant studies should be included (essential). If the bibliography includes unpublished studies, copies of these studies, or adequate abstracts or summaries, should be attached (essential). Where revisions have been made in a selection procedure to assure compatability between successful job performance and the probability of being selected, the studies underlying such revisions should be included (essential). All statistical results should be organized and presented by relevant race, sex, and ethnic group (essential).

(9) *Alternative procedures investigated.* The selection procedures investigated and available evidence of their impact should be identified (essential). The scope, method, and findings of investigation, and the conclusions reached in light of the findings, should be fully described (essential).

(10) *Uses and applications.* The methods considered for use of the selection procedure (e.g., as a screening device with a cutoff score, for grouping or ranking, or combined with other

procedures in a battery) and available evidence of their impact should be described (essential). This description should include the rationale for choosing the method for operational use, and the evidence of the validity and utility of the procedure as it is to be used (essential). The purpose for which the procedure is to be used (e.g., hiring, transfer, promotion) should be described (essential). If weights are assigned to different parts of the selection procedure, these weights and the validity of the weighted composite should be reported (essential). If the selection procedure is used with a cutoff score, the user should describe the way in which normal expectations of proficiency within the work force were determined and the way in which the cutoff score was determined (essential).

(11) *Source data.* Each user should maintain records showing all pertinent information about individual sample members and raters where they are used, in studies involving the validation of selection procedures. These records should be made available upon request of a compliance agency. In the case of individual sample members these data should include scores on the selection procedure(s), scores on criterion measures, age, sex, race, or ethnic group status, and experience on the specific job on which the validation study was conducted, and may also include such things as education, training, and prior job experience, but should not include names and social security numbers. Records should be maintained which show the ratings given to each sample member by each rater.

(12) *Contact person.* The name, mailing address, and telephone number of the person who may be contacted for further information about the validity study should be provided (essential).

(13) *Accuracy and completeness.* The report should describe the steps taken to assure the accuracy and completeness of the collection, analysis, and report of data and results.

C. *Content validity studies.* Reports of content validity for a selection procedure should include the following information:

(1) *User(s), location(s) and date(s) of study.* Dates and location(s) of the job analysis should be shown (essential).

(2) *Problem and setting.* An explicit definition of the purpose(s) of the study and the circumstances in which the study was conducted should be provided. A description of existing selection procedures and cutoff scores, if any, should be provided.

(3) *Job analysis—Content of the job.* A description of the method used to analyze the job should be provided (essential). The work behavior(s), the associated tasks, and, if the behavior results in a work product, the work products should be completely described (essential). Measures of criticality and/or importance of the work behavior(s) and the method of determining these measures should be provided (essential). Where the job analysis also identified the knowledges, skills, and abilities used in work behavior(s), an operational definition for each knowledge in terms of a body of learned information and for each skill and ability in terms of observable behaviors and outcomes, and the relationship between each knowledge, skill, or ability and each work behavior, as well as the method used to determine this relationship, should be provided (essential). The work situation should be described, including the setting in which work behavior(s) are performed, and where appropriate, the manner in which knowledges, skills, or abilities are used, and the complexity and difficulty of the knowledge, skill, or ability as used in the work behavior(s).

(4) *Selection procedure and its content.* Selection procedures, including those constructed by or for the user, specific training requirements, composites of selection procedures, and any other procedure supported by content validity, should be completely and explicitly described or attached (essential). If commercially available selection procedures are used, they should be

described by title, form, and publisher (essential). The behaviors measured or sampled by the selection procedure should be explicitly described (essential). Where the selection procedure purports to measure a knowledge, skill, or ability, evidence that the selection procedure measures and is a representative sample of the knowledge, skill, or ability should be provided (essential).

(5) *Relationship between the selection procedure and the job.* The evidence demonstrating that the selection procedure is a representative work sample, a representative sample of the work behavior(s), or a representative sample of a knowledge, skill, or ability as used as a part of a work behavior and necessary for that behavior should be provided (essential). The user should identify the work behavior(s) which each item or part of the selection procedure is intended to sample or measure (essential). Where the selection procedure purports to sample a work behavior or to provide a sample of a work product, a comparison should be provided of the manner, setting, and the level of complexity of the selection procedure with those of the work situation (essential). If any steps were taken to reduce adverse impact on a race, sex, or ethnic group in the content of the procedure or in its administration, these steps should be described. Establishment of time limits, if any, and how these limits are related to the speed with which duties must be performed on the job, should be explained. Measures of central tendency (e.g., means) and measures of dispersion (e.g., standard deviations) and estimates of realibility should be reported for all selection procedures if available. Such reports should be made for relevant race, sex, and ethnic subgroups, at least on a statistically reliable sample basis.

(6) *Alternative procedures investigated.* The alternative selection procedures investigated and available evidence of their impact should be identified (essential). The scope, method, and findings of the investigation, and the conclusions reached in light of the findings, should be fully described (essential).

(7) *Uses and applications.* The methods considered for use of the selection procedure (e.g., as a screening device with a cutoff score, for grouping or ranking, or combined with other procedures in a battery) and available evidence of their impact should be described (essential). This description should include the rationale for choosing the method for operational use, and the evidence of the validity and utility of the procedure as it is to be used (essential). The purpose for which the procedure is to be used (e.g., hiring, transfer, promotion) should be described (essential). If the selection procedure is used with a cutoff score, the user should describe the way in which normal expectations of proficiency within the work force were determined and the way in which the cutoff score was determined (essential). In addition, if the selection procedure is to be used for ranking, the user should specify the evidence showing that a higher score on the selection procedure is likely to result in better job performance.

(8) *Contact person.* The name, mailing address, and telephone number of the person who may be contacted for further information about the validity study should be provided (essential).

(9) *Accuracy and completeness.* The report should describe the steps taken to assure the accuracy and completeness of the collection, analysis, and report of data and results.

D. *Construct validity studies.* Reports of construct validity for a selection procedure should include the following information:

(1) *User(s), location(s), and date(s) of study.* Date(s) and location(s) of the job analysis and the gathering of other evidence called for by these guidelines should be provided (essential).

(2) *Problem and setting.* An explicit definition of the purpose(s) of the study and the circumstances in which the study was conducted should be provided. A description of existing

selection procedures and cutoff scores, if any, should be provided.

(3) *Construct definition.* A clear definition of the construct(s) which are believed to underlie successful performance of the critical or important work behavior(s) should be provided (essential). This definition should include the levels of construct performance relevant to the job(s) for which the selection procedure is to be used (essential). There should be a summary of the position of the construct in the psychological literature, or in the absence of such a position, a description of the way in which the definition and measurement of the construct was developed and the psychological theory underlying it (essential). Any quantitative data which identify or define the job constructs, such as factor analyses, should be provided (essential).

(4) *Job analysis.* A description of the method used to analyze the job should be provided (essential). A complete description of the work behavior(s) and, to the extent appropriate, work outcomes and measures of their criticality and/or importance should be provided (essential). The report should also describe the basis on which the behavior(s) or outcomes were determined to be important, such as their level of difficulty, their frequency of performance, the consequences of error or other appropriate factors (essential). Where jobs are grouped or compared for the purposes of generalizing validity evidence, the work behavior(s) and work product(s) for each of the jobs should be described, and conclusions concerning the similarity of the jobs in terms of observable work behaviors or work products should be made (essential).

(5) *Job titles and codes.* It is desirable to provide the selection procedure user's job title(s) for the job(s) in question and the corresponding job title(s) and code(s) from the United States Employment Service's dictionary of occupational titles.

(6) *Selection procedure.* The selection procedure used as a measure of the construct should be completely and explicitly

described or attached (essential). If commercially available selection procedures are used, they should be identified by title, form and publisher (essential). The research evidence of the relationship between the selection procedure and the construct, such as factor structure, should be included (essential). Measures of central tendency, variability and reliability of the selection procedure should be provided (essential). Whenever feasible, these measures should be provided separately for each relevant race, sex and ethnic group.

(7) *Relationship to job performance.* The criterion-related study(ies) and other empirical evidence of the relationship between the construct measured by the selection procedure and the related work behavior(s) for the job or jobs in question should be provided (essential). Documentation of the criterion-related study(ies) should satisfy the provisions of section 15B above or section 15E(1) below, except for studies conducted prior to the effective date of these guidelines (essential). Where a study pertains to a group of jobs, and, on the basis of the study, validity is asserted for a job in the group, the observed work behaviors and the observed work products for each of the jobs should be described (essential). Any other evidence used in determining whether the work behavior(s) in each of the jobs is the same should be fully described (essential).

(8) *Alternative procedures investigated.* The alternative selection procedures investigated and available evidence of their impact should be identified (essential). The scope, method, and findings of the investigation, and the conclusions reached in light of the findings should be fully described (essential).

(9) *Uses and applications.* The methods considered for use of the selection procedure (e.g., as a screening device with a cutoff score, for grouping or ranking, or combined with other procedures in a battery) and available evidence of their impact should be described (essential). This description should include the rationale for choosing the method for operational use and

the evidence of the validity and utility of the procedure as it is to be used (essential). The purpose for which the procedure is to be used (e.g., hiring, transfer, promotion) should be described (essential). If weights are assigned to different parts of the selection procedure, these weights and the validity of the weighted composite should be reported (essential). If the selection procedure is used with a cutoff score, the user should describe the way in which normal expectations of proficiency within the work force were determined and the way in which the cutoff score was determined (essential).

(10) *Accuracy and completeness.* The report should describe the steps taken to assure the accuracy and completeness of the collection, analysis, and report of data and results.

(11) *Source data.* Each user should maintain records showing all pertinent information relating to its study of construct validity.

(12) *Contact person.* The name, mailing address, and telephone number of the individual who may be contacted for further information about the validity study should be provided (essential).

E. *Evidence of validity from other studies.* When validity of a selection procedure is supported by studies not done by the user, the evidence from the original study or studies should be compiled in a manner similar to that required in the appropriate section of this section 15 above. In addition, the following evidence should be supplied:

(1) *Evidence from criterion-related validity studies.*—a. *Job information.* A description of the important job behavior(s) of the user's job and the basis on which the behaviors were determined to be important should be provided (essential). A full description of the basis for determining that these important work behaviors are the same as those of the job in the original study (or studies) should be provided (essential).

b. *Relevance of criteria.* A full description of the basis on which the criteria used in the original studies are determined to be relevant for the user should be provided (essential).

c. *Other variables.* The similarity of important applicant pool or sample characteristics reported in the original studies to those of the user should be described (essential). A description of the comparison between the race, sex and ethnic composition of the user's relevant labor market and the sample in the original validity studies should be provided (essential).

d. *Use of the selection procedure.* A full description should be provided showing that the use to be made of the selection procedure is consistent with the findings of the original validity studies (essential).

e. *Bibliography.* A bibliography of reports of validity of the selection procedure for the job or jobs in question should be provided (essential). Where any of the studies included an investigation of test fairness, the results of this investigation should be provided (essential). Copies of reports published in journals that are not commonly available should be described in detail or attached (essential). Where a user is relying upon unpublished studies, a reasonable effort should be made to obtain these studies. If these unpublished studies are the sole source of validity evidence they should be described in detail or attached (essential). If these studies are not available, the name and address of the source, an adequate abstract or summary of the validity study and data, and a contact person in the source organization should be provided (essential).

(2) *Evidence from content validity studies.* See section 14C(3) and section 15C above.

(3) *Evidence from construct validity studies.* See sections 14D(2) and 15D above.

F. *Evidence of validity from cooperative studies.* Where a selection procedure has been validated through a cooperative

study, evidence that the study satisfies the requirements of sections 7, 8 and 15E should be provided (essential).

G. *Selection for higher level job.* If a selection procedure is used to evaluate candidates for jobs at a higher level than those for which they will initially be employed, the validity evidence should satisfy the documentation provisions of this section 15 for the higher level job or jobs, and in addition, the user should provide: (1) a description of the job progression structure, formal or informal; (2) the data showing how many employees progress to the higher level job and the length of time needed to make this progression; and (3) an identification of any anticipated changes in the higher level job. In addition, if the test measures a knowledge, skill or ability, the user should provide evidence that the knowledge, skill or ability is required for the higher level job and the basis for the conclusion that the knowledge, skill or ability is not expected to develop from the training or experience on the job.

H. *Interim use of selection procedures.* If a selection procedure is being used on an interim basis because the procedure is not fully supported by the required evidence of validity, the user should maintain and have available (1) substantial evidence of validity for the procedure, and (2) a report showing the date on which the study to gather the additional evidence commenced, the estimated completion date of the study, and a description of the data to be collected (essential).

DEFINITIONS

§ 1607.16 Definitions.

The following definitions shall apply throughout these guidelines:

A. *Ability.* A present competence to perform an observable behavior or a behavior which results in an observable product.

B. *Adverse impact.* A substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group. See section 4 of these guidelines.

C. *Compliance with these guidelines.* Use of a selection procedure is in compliance with these guidelines if such use has been validated in accord with these guidelines (as defined below), or if such use does not result in adverse impact on any race, sex, or ethnic group (see section 4, above), or, in unusual circumstances, if use of the procedure is otherwise justified in accord with Federal law. See section 6B, above.

D. *Content validity.* Demonstrated by data showing that the content of a selection procedure is representative of important aspects of performance on the job. See section 5B and section 14C.

E. *Construct validity.* Demonstrated by data showing that the selection procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important for successful job performance. See section 5B and section 14D.

F. *Criterion-related validity.* Demonstrated by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior. See sections 5B and 14B.

G. *Employer.* Any employer subject to the provisions of the Civil Rights Act of 1964, as amended, including State or local governments and any Federal agency subject to the provisions of section 717 of the Civil Rights Act of 1964, as amended, and any Federal contractor or subcontractor or federally assisted construction contractor or subcontractor covered by Executive Order 11246, as amended.

H. *Employment agency.* Any employment agency subject to the provisions of the Civil Rights Act of 1964, as amended.

I. *Enforcement action.* For the purposes of section 4 a proceeding by a Federal enforcement agency such as a lawsuit or an administrative proceeding leading to debarment from or withholding, suspension, or termination of Federal Government contracts or the suspension or withholding of Federal Government funds; but not a finding of reasonable cause or a conciliation process or the issuance of right to sue letters under title VII or under Executive Order 11246 where such finding, conciliation, or issuance of notice of right to sue is based upon an individual complaint.

J. *Enforcement agency.* Any agency of the executive branch of the Federal Government which adopts these guidelines for purposes of the enforcement of the equal employment opportunity laws or which has responsibility for securing compliance with them.

K. *Job analysis.* A detailed statement of work behaviors and other information relevant to the job.

L. *Job description.* A general statement of job duties and responsibilities.

M. *Knowledge.* A body of information applied directly to the performance of a function.

N. *Labor organization.* Any labor organization subject to the provisions of the Civil Rights Act of 1964, as amended, and any committee subject thereto controlling apprenticeship or other training.

O. *Observable.* Able to be seen, heard, or otherwise perceived by a person other than the person performing the action.

P. *Race, sex, or ethnic group.* Any group of persons identifiable on the grounds of race, color, religion, sex, or national origin.

Q. *Selection procedure.* Any measure, combination of measures, or procedure used as a basis for any employment

decision. Selection procedures include the full range of assessment techniques from traditional paper and pencil tests, performance tests, training programs, or probationary periods and physical, educational, and work experience requirements through informal or casual interviews and unscored application forms.

R. *Selection rate.* The proportion of applicants or candidates who are hired, promoted, or otherwise selected.

S. *Should.* The term "should" as used in these guidelines is intended to connote action which is necessary to achieve compliance with the guidelines, while recognizing that there are circumstances where alternative courses of action are open to users.

T. *Skill.* A present, observable competence to perform a learned psychomoter act.

U. *Technical feasibility.* The existence of conditions permitting the conduct of meaningful criterion-related validity studies. These conditions include: (1) An adequate sample of persons available for the study to achieve findings of statistical significance; (2) having or being able to obtain a sufficient range of scores on the selection procedure and job performance measures to produce validity results which can be expected to be representative of the results if the ranges normally expected were utilized; and (3) having or being able to devise unbiased, reliable and relevant measures of job performance or other criteria of employee adequacy. See section 14B(2). With respect to investigation of possible unfairness, the same considerations are applicable to each group for which the study is made. See section 14B(8).

V. *Unfairness of selection procedure.* A condition in which members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences are not reflected in differences in measures of job performance. See section 14B(7).

W. *User.* Any employers, labor organization, employment agency, or licensing or certification board, to the extent it may be covered by Federal equal employment opportunity law, which uses a selection procedure as a basis for any employment decision. Whenever an employer, labor organization, or employment agency is required by law to restrict recruitment for any occupation to those applicants who have met licensing or certification requirements, the licensing or certifying authority to the extent it may be covered by Federal equal employment opportunity law will be considered the user with respect to those licensing or certification requirements. Whenever a State employment agency or service does no more than administer or monitor a procedure as permitted by Department of Labor regulations, and does so without making referrals or taking any other action on the basis of the results, the State employment agency will not be deemed to be a user.

X. *Validated in accord with these guidelines or properly validated.* A demonstration that one or more validity study or studies meeting the standards of these guidelines has been conducted, including investigation and, where appropriate, use of suitable alternative selection procedures as contemplated by section 3B, and has produced evidence of validity sufficient to warrant use of the procedure for the intended purpose under the standards of these guidelines.

Y. *Work behavior.* An activity performed to achieve the objectives of the job. Work behaviors involve observable (physical) components and unobservable (mental) components. A work behavior consists of the performance of one or more tasks. Knowledges, skills, and abilities are not behaviors, although they may be applied in work behaviors.

APPENDIX

§ 1607.17 Policy statement on affirmative action (see section 13B).

The Equal Employment Opportunity Coordinating Council was established by act of Congress in 1972, and charged with responsibility for developing and implementing agreements and policies designed, among other things, to eliminate conflict and inconsistency among the agencies of the Federal Government responsible for administering Federal law prohibiting discrimination on grounds of race, color, sex, religion, and national origin. This statement is issued as an initial response to the requests of a number of State and local officials for clarification of the Government's policies concerning the role of affirmative action in the overall equal employment opportunity program. While the Coordinating Council's adoption of this statement expresses only the views of the signatory agencies concerning this important subject, the principles set forth below should serve as policy guidance for other Federal agencies as well.

(1) Equal employment opportunity is the law of the land. In the public sector of our society this means that all persons, regardless of race, color, religion, sex, or national origin shall have equal access to positions in the public service limited only by their ability to do the job. There is ample evidence in all sectors of our society that such equal access frequently has been denied to members of certain groups because of their sex, racial, or ethnic characteristics. The remedy for such past and present discrimination is twofold.

On the one hand, vigorous enforcement of the laws against discrimination is essential. But equally, and perhaps even more important are affirmative, voluntary efforts on the part of public employers to assure that positions in the public service are genuinely and equally accessible to qualified persons, without regard to their sex, racial, or ethnic characteristics.

Without such efforts equal employment opportunity is no more than a wish. The importance of voluntary affirmative action on the part of employers is underscored by title VII of the Civil Rights Act of 1964, Executive Order 11246, and related laws and regulations—all of which emphasize voluntary action to achieve equal employment opportunity.

As with most management objectives, a systematic plan based on sound organizational analysis and problem identification is crucial to the accomplishment of affirmative action objectives. For this reason, the Council urges all State and local governments to develop and implement results oriented affirmative action plans which deal with the problems so identified.

The following paragraphs are intended to assist State and local governments by illustrating the kinds of analyses and activities which may be appropriate for a public employer's voluntary affirmative action plan. This statement does not address remedies imposed after a finding of unlawful discrimination.

(2) Voluntary affirmative action to assure equal employment opportunity is appropriate at any stage of the employment process. The first step in the construction of any affirmative action plan should be an analysis of the employer's work force to determine whether precentages of sex, race, or ethnic groups in individual job classifications are substantially similar to the percentages of those groups available in the relevant job market who possess the basic job-related qualifications.

When substantial disparities are found through such analyses, each element of the overall selection process should be examined to determine which elements operate to exclude persons on the basis of sex, race, or ethnic group. Such elements include, but are not limited to, recruitment, testing, ranking certification, interview, recommendations for selection, hiring, promotion, etc. The examination of each element of the selection process should at a minimum include a determination of its validity in predicting job performance.

(3) When an employer has reason to believe that its selection procedures have the exclusionary effect described in paragraph 2 above, it should initiate affirmative steps to remedy the situation. Such steps, which in design and execution may be race, color, sex, or ethnic "conscious," include, but are not limited to, the following:

(a) The establishment of a long-term goal, and short-range, interim goals and timetables for the specific job classifications, all of which should take into account the availability of basically qualified persons in the relevant job market;

(b) A recruitment program designed to attract qualified members of the group in question;

(c) A systematic effort to organize work and redesign jobs in ways that provide opportunities for persons lacking "journeyman" level knowledge or skills to enter and, with appropriate training, to progress in a career field;

(d) Revamping selection instruments or procedures which have not yet been validated in order to reduce or eliminate exclusionary effects on particular groups in particular job classifications;

(e) The initiation of measures designed to assure that members of the affected group who are qualified to perform the job are included within the pool of persons from which the selecting official makes the selection;

(f) A systematic effort to provide career advancement training, both classroom and on-the-job, to employees locked into dead end jobs; and

(g) The establishment of a system for regularly monitoring the effectiveness of the particular affirmative action program, and procedures for making timely adjustments in this program where effectiveness is not demonstrated.

(4) The goal of any affirmative action plan should be achievement of genuine equal employment opportunity for all qualified persons. Selection under such plans should be based upon the ability of the applicant(s) to do the work. Such plans should not require the selection of the unqualified, or the unneeded, nor should they require the selection of persons on the basis of race, color, sex, religion, or national origin. Moreover, while the Council believes that this statement should serve to assist State and local employers, a well as Federal agencies, it recognizes that affirmative action cannot be viewed as a standardized program which must be accomplished in the same way at all times in all places.

Accordingly, the Council has not attempted to set forth here either the minimum or maximum voluntary steps that employers may take to deal with their respective situations. Rather, the Council recognizes that under applicable authorities, State and local employers have flexibility to formulate affirmative action plans that are best suited to their particular situations. In this manner, the Council believes that affirmative action programs will best serve the goal of equal employment opportunity.

Respectfully submitted,

Harold R. Tyler, Jr.,
Deputy Attorney General and Chairman of the Equal Employment Coordinating Council.

Michael H. Moskow,
Under Secretary of Labor.

Ethel Bent Walsh,
Acting Chairman, Equal Employment Opportunity Commission.

Robert E. Hampton,
Chairman, Civil Service Commission.

Arthur E. Flemming,
Chairman, Commission on Civil Rights.

Because of its equal employment opportunity responsibilities under the State and Local Government Fiscal Assistance Act of 1972 (the revenue sharing act), the Department of Treasury was invited to participate in the formulation of this policy statement; and it concurs and joins in the adoption of this policy statement.

Done this 26th day of August 1976.

Richard Albrecht,
General Counsel,
Department of the Treasury.

§ 1607.18 Citations.

The official title of these guidelines is "Uniform Guidelines on Employee Selection Procedures (1978)". The Uniform Guidelines on Employee Selection Procedures (1978) are intended to establish a uniform Federal position in the area of prohibiting discrimination in employment practices on grounds of race, color, religion, sex, or national origin. These guidelines have been adopted by the Equal Employment Opportunity Commission, the Department of Labor, the Department of Justice, and the Civil Service Commission.

The official citation is:

"Section ——, Uniform Guidelines on Employee Selection Procedure (1978); 43 FR —— (August 25, 1978)."

The short form citation is:

"Section ——, U.G.E.S.P. (1978); 43 FR —— (August 25, 1978)."

When the guidelines are cited in connection with the activities of one of the issuing agencies, a specific citation to the regulations of that agency can be added at the end of the above citation. The specific additional citations are as follows:

Equal Employment Opportunity Commission
29 CFR Part 1607
Department of Labor
Office of Federal Contract Compliance Programs
41 CFR Part 60-3
Department of Justice
28 CFR 50.14
Civil Service Commission
5 CFR 300.103(c)

Normally when citing these guidelines, the section number immediately preceding the title of the guidelines will be from these guidelines series 1-18. If a section number from the codification for an individual agency is needed it can also be added at the end of the agency citation. For example, section 6A of these guidelines could be cited for EEOC as follows: "Section 6A, Uniform Guidelines on Employee Selection Procedures (1978); 43 FR ——, (August 25, 1978); 29 CFR Part 1607, section 6A."

APPENDIX G

Department of Justice
Guidelines on Employee Selection Procedures.
28 C.F.R. §50.14 (1976)

The guidelines set forth below are intended as a statement of policy of the Department of Justice and will be applied by the Department in exercising its responsibilities under Federal law relating to equal employment opportunity.

PART I—GENERAL PRINCIPLES

§1 Statement of Purpose

These guidelines are intended to be a set of principles which will assist employers, labor organizations, employment agencies, and licensing and certification boards in complying with equal employment opportunity requirements of Federal law with respect to race, color, religion, sex and national origin. They are designed to provide a framework for determining the proper use of tests and other selection procedures consistent with Federal law. These guidelines do not require a user to conduct validity studies of selection procedures where no adverse impact results. However, all users are encouraged to use selection procedures which are valid, especially users operating under merit principles. Nothing in these guidelines is intended or should be interpreted as discouraging the use of procedures which have been properly validated in accordance with these guidelines for the purpose of determining qualifications or selecting on the basis of relative qualifications. Nothing in these guidelines is intended to apply to persons not subject to the requirements of Title VII, Executive Order 11246, or other equal employment opportunity requirements of Federal law. These guidelines are not intended to apply to any responsibilities an employer, employment agency labor organization may have under the Age Discrimination Act of 1975 not to discriminate

on the basis of age, or under section 504 of the Rehabilitation Act of 1973 not to discriminate on the basis of handicap. Nothing contained in these guidelines is intended to interfere with any obligation imposed or right granted by Federal law to users to extend a publicly announced preference in employment to Indians living on or near an Indian reservation in connection with employment opportunities on or near an Indian reservation.

§2 Scope

a. These guidelines will be applied by the Department of Labor to contractors and subcontractors subject to Executive Order 11246 as amended by Executive Order 11375 (hereinafter "Executive Order 11246"); and by the Civil Service Commission to federal agencies subject to Sec. 717 of the Civil Rights Act of 1964, as amended by the Equal Employment Opportunity Act of 1972 (hereinafter "the Civil Rights Act of 1964") and to its responsibilities toward state and local governments under Section 208(b)(1) of the Intergovernmental Personnel Act; by the Department of Justice in exercising its responsibilities under Federal law; and by any other Federal agency which adopts them.

b. These guidelines apply to selection procedures which are used as a basis for any employment decision. Employment decisions include but are not limited to hire, promotion, demotion, membership (for example in a labor organization), referral, retention, licensing and certification, to the extent that licensing and certification may be covered by Federal equal employment opportunity law. Selection for training is also considered an employment decision if it leads to any of the decisions listed above.

c. These guidelines do not apply to the use of a bona fide seniority system within the meaning of Title VII of the Civil Rights Act of 1964, as amended, as defined by Federal appellate

court decisions, for any employment decision. These guidelines do not call for the validation of such a seniority system used as a basis for such employment decision and the use of such a seniority system as a basis for such employment decisions is consistent with these guidelines.

d. These guidelines do not apply to the entire range of Federal equal employment opportunity law, but only to selection procedures which are used as a basis for making employment decisions. For example, the use of recruiting procedures designed to attract racial, ethnic or sex groups which were previously denied employment opportunities or which are presently under-utilized may be necessary to bring an employer into compliance with Federal law, and is frequently an essential element to any effective affirmative action program; but the subject of recruitment practices is not addressed by these guidelines because that subject concerns procedures other than selection procedures.

§3 Relationship Between Use of Selection Procedures and Discrimination

a. The use of any selection procedure which has an adverse impact on the members of any racial, ethnic or sex group with respect to any employment decision will be considered to be discriminatory and inconsistent with these guidelines, unless the procedure is validated in accordance with the principles contained in these guidelines or unless use of the procedure is warranted under § 3b.

b. There are circumstances in which it is not feasible or not appropriate to utilize the validation techniques contemplated by these guidelines. In such circumstances, the user should utilize selection procedures which are as job related as possible and which will minimize or eliminate adverse impact. (1) When an unstandardized, informal or unscored selection procedure which has an adverse impact is utilized, the user should seek in-

sofar as possible to eliminate the adverse impact, or, if feasible, to modify the procedure to one which is a formal, scored or qualified measure or combination of measures and then to validate the procedure in accord with these guidelines, or otherwise to justify continued use of the procedure in accord with Federal law. (2) When a standardized, formal or scored selection procedure is used for which it is not feasible or not appropriate to utilize the validation techniques contemplated by these guidelines, the user should either modify the procedure to eliminate the adverse impact or otherwise justify continued use of the procedure in accord with Federal law.

c. Generally where alternative selection procedures are available which have been shown to be equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact. Accordingly, whenever a validity study is called for by these guidelines, the user should make a reasonable effort to investigate suitable alternative selection procedures which have as little adverse impact as possible, for the purpose of determining the appropriateness of using or validating them in accord with these guidelines. If a user has made a reasonable effort to become aware of such alternative procedures and a validity study for a job or group of jobs has been made in accord with these guidelines, the use of the selection procedure may continue until such time as it should reasonably be reviewed for currency. Whenever the user is shown a suitable alternative selection procedure with evidence of at least equal validity and less adverse impact, the user should investigate if the purpose of determining the appropriateness of using or validating it in accord with these guidelines. This subsection is not intended to preclude the combination of procedures into a significantly more valid procedure, if such a combination has been properly validated.

§4 Information on Impact

a. Each user should have available for inspection records or other information which will disclose the impact which its selec-

tion procedures have upon employment opportunities of persons by identifiable racial, ethnic or sex groups in order to determine compliance with the provisions of § 3 above. Where there are large numbers of applicants and procedures are administered frequently, such information may be retained on a sample basis provided that the sample is appropriate in terms of the applicant population and adequate in size. The records called for by this section are to be maintained by sex, and by racial and ethnic groups as follows: blacks (Negroes), American Indians (including Alaskan Natives), Asians (including Pacific islanders), Hispanic (including persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or culture regardless of race), whites (Caucasians) other than Hispanic and totals. The classifications called for by this section are intended to be consistent with the Employer Information (EEO-1 et seq.) series of reports. The user should adopt safeguards to insure that records of race, color, religion, sex, or national origin are used for appropriate purposes such as determining adverse impact, or (where required) for developing and monitoring affirmative action programs, and that such records are not used for making employment decisions.

b. The information called for by this section should be examined for possible diverse impact. If the records called for by this section indicate that the total selection process for a job has no adverse impact, the individual components of the selection process need not be evaluated separately for adverse impact. If a total selection process does have adverse impact, the individual components of the selection process should be evaluated for diverse impact.

A selection rate for any racial, ethnic or sex group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded as evidence of adverse impact. Smaller differences in selection rate may nevertheless be con-

sidered to constitute diverse impact, where they are significant in both statistical and practical terms. Greater differences in selection rate would not necessarily be regarded as constituting adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group.

c. Federal agencies which adopt these guidelines for purposes of the enforcement of the equal employment opportunity laws or which have responsibility for securing compliance with them (hereafter referred to as enforcement agencies) will consider in carrying out their obligations the general posture of the user with respect to equal employment opportunity for the job classification or group of classifications in question. Where a user has adopted an affirmative action program, the Federal enforcement agencies will consider the provisions of that program, including the goals and timetables which the employer has adopted and the progress which the employer has made in carrying out that program and in meeting the goals and timetables. These guidelines recognize that a user is prohibited by Federal law from the making of employment decisions on the basis of race and color and (except for bona fide occupational qualifications) on the basis of sex, religion and national origin; and nothing in this subsection or in the guidelines is intended to encourage or permit the granting of preferential treatment to any individual or to any group because of the race, color, religion, sex or national origin of such individual or group.

§5 General Standards for Validity Studies

a. For the purposes of satisfying these guidelines users may rely upon criterion related validity studies, content validity studies or construct validity studies, in accordance with the standards set forth in Part II of these guidelines, § 12 infra.

b. These guidelines are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other assessment techniques, such as those described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (American Psychological Association, Washington, D.C. 1974) (hereinafter "APA Standards"), and standard text books and journals in the field of personnel selection.

c. For any selection procedure which has an adverse impact each user should maintain and have available such documentation as is described in Part III of these guidelines, § 13 infra.

d. Selection procedures subject to validity studies under § 3a above should be administered and scored under standardized conditions.

e. In general, users should avoid making employment decisions on the basis of measures of knowledges, skills or abilities which are normally learned in a brief orientation period, and which have an adverse impact.

f. Where cut off scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force. Where other factors are used in determining cut off scores, such as the relationship between the number of vacancies and the number of applicants, the degree of adverse impact should be considered.

g. Selection procedures may be used to predict the performance of candidates for a job which is at a higher level than the job for which the person is initially being selected if a majority of the individuals who remain employed will progress to the higher level within a reasonable period of time. A "reasonable period of time" will vary for different jobs and employment situations but will seldom be more than five years. Examining

for a higher level job would not be appropriate (1) if the majority of those remaining employed do not progress to the higher level job, (2) if there is a reason to doubt that the higher level job will continue to require essentially similar skills during the progression period, or (3) if knowledges, skills or abilities required for advancement would be expected to develop principally from the training or experience on the job.

h. Users may continue the use of a selection procedure which is not at the moment fully supported by the required evidence of validity, provided: (1) the user can cite substantial evidence of validity in accord with these guidelines and (2) the user has in progress when technically feasible, studies which are designed to produce the additional data required within a reasonable time.

If the additional studies do not produce the data required to demonstrate validity, the user is not relieved of or protected against any obligations arising under federal law.

i. Whenever a validity study has been made in accord with these guidelines for the use of a particular selection procedure for a job or group of jobs, additional studies need not be performed until such time as the validity study is subject to review as provided in § 3c above. There are no absolutes in the area of determining the currency of a validity study. All circumstances concerning the study, including the validation strategy used, and changes in the relevant labor market and the job should be considered in the determination of when a validity study is outdated.

§6 Cooperative Validity Studies and Use of Other Validity Studies

a. It is the intent of the agencies issuing these guidelines to encourage and facilitate cooperative development and validation efforts by employers, labor organizations and employment agencies to achieve selection procedures which are consistent with these guidelines.

b. Criterion-related validity studies conducted by one test user, or described in test manuals and the professional literature, will be considered acceptable for use by another user when: (1) the weight of the evidence from studies meeting the standards of § 12b below shows that the selection procedure is valid; (2) the studies pertain to a job which has substantially the same major job duties as shown by appropriate job analyses and (3) the studies include a study of test fairness for those racial, ethnic and sex subgroups which constitute significant factors in the borrowing user's relevant labor market for the job or jobs in question. If the studies relied upon satisfy (1) and (2) above but do not contain an investigation of test fairness, and it is not technically feasible for the borrowing user to conduct an internal study of test fairness, the borrowing user may utilize the study until studies conducted elsewhere show test unfairness, or until such time as it becomes technically feasible to conduct an internal study of test fairness and the results of that study can be acted upon.

If it is technically feasible for a borrowing user to conduct an internal validity study, and there are variables in the other studies which are likely to affect validity or fairness significantly, the user may rely upon such studies only on an interim basis in accord with § 5h, and will be expected to conduct an internal validity study in accord with § 12b below. Otherwise the borrowing user may rely upon such acceptable studies for operational use without an internal study.

c. Selection procedures shown by one user to be content valid in accord with § 12c will be considered acceptable for use by another user for a performance domain if the borrowing user's job analysis shows that the same performance domain is present in the borrowing user's job. The selection procedure may be used operationally if the conditions of § 12c(3) and § 12c(6) are satisfied by the borrowing user.

d. The conditions under which findings of construct validity may be generalized are described in § 12d(4).

e. If validity evidence from a multiunit or cooperative study satisfies the requirements of subparagraphs b, c or d above, evidence of validity specific to each unit or user usually will not be required unless there are variables in the units not studied which are likely to affect validity significantly.

§7 No Assumption of Validity

a. Under no circumstances will the general reputation of a selection procedure, its author or its publisher or casual reports of its validity be accepted in lieu of evidence of validity. Specifically ruled out are: Assumptions of validity based on a procedure's name or descriptive labels; all forms of promotional literature; data bearing on the frequency of a procedure's usage; testimonial statements and credentials of sellers, users, or consultants; and other non-empirical or anecdotal accounts of selection practices or selection outcomes.

b. Professional supervision of selection activities is encouraged but is not a substitute for documented evidence of validity. The enforcement agencies will take into account the fact that a thorough job analysis and careful development of a selection procedure enhances the probability that the selection procedure is valid for the job.

§8 Employment Agencies and Employment Services

a. An employment agency, including private employment agencies and State employment agencies, which agrees to a request by an employer or labor organization to devise and utilize a selection procedure should follow the standards for determining adverse impact and, if adverse impact is demonstrated, show validity as set forth in these guidelines. An employment agency is not relieved of its obligation herein because the user did not request such validation or has requested the use of some lesser standard of validation than is provided in these guidelines. The

use of an employment agency does not relieve an employer or labor organization of its responsibilities under Federal law to provide equal employment opportunity or its obligations as a user under these guidelines.

b. Where an employment agency or service is requested to administer a selection program which has been devised elsewhere and to make referrals pursuant to the results, the employment agency or service should obtain evidence of the absence of adverse impact, or of validity, as described in these guidelines, before it administers the selection program and makes referrals pursuant to the results. The employment agency must furnish on request such evidence of validity. An employment agency or service will be expected to refuse to make referrals based on the selection procedure where the employer or labor organization does not supply satisfactory evidence of validity or lack of adverse impact.

§9 Disparate Treatment

The principle of disparate or unequal treatment must be distinguished from the concepts of validation. A selection procedure—even though validated against job performance in accordance with the guidelines in this part—cannot be imposed upon members of a racial, sex or ethnic group where other employees, applicants, or members have not been subjected to that standard. Disparate treatment occurs where members of a racial, sex, or ethnic group have been denied the same employment, promotion, transfer or membership opportunities as have been made available to other employees or applicants. Those employees or applicants who have been denied equal treatment, because of prior discriminatory practices or policies, must at least be afforded the same opportunities as had existed for other employees or applicants during the period of discrimination. Thus, the persons who were in the class of persons discriminated against and were available in the relevant job market during the period the user followed the discriminatory practices should be

allowed the opportunity to qualify under the less stringent selection procedures previously followed, unless the user demonstrates that the increased standards are required for the safety or efficiency of the operation. Nothing in this section is intended to prohibit a user who has not previously followed merit standards from adopting merit standards; nor does it preclude a user who has previously used invalid or unvalidated selection procedures from developing and using procedures which are validated in accord with these guidelines.

§10 Retesting

Users should provide a reasonable opportunity for retesting and reconsideration. The user may however take reasonable steps to preserve the security of its procedures. Where examinations are administered periodically with public notice, such reasonable opportunity exists, unless persons who have previously been tested are precluded from retesting.

§11 Affirmative Action

The use of selection procedures which have been validated pursuant to these guidelines does not relieve users of any obligations they may have to undertake affirmative action to assure equal employment opportunity. Nothing in these guidelines is intended to preclude the use of selection procedures (consistent with Federal law—see § 4c) which assist in the achievement of affirmative action objectives.

PART II—TECHNICAL STANDARDS

§12 Technical Standards for Validity Studies

The following minimum standards, as applicable, should be met in conducting a validity study. Nothing in these guidelines is intended to preclude the development and use of other profes-

sionally acceptable techniques with respect to validation of selection procedures.

a. Any validity study should be based upon a review of information about the job for which the selection procedure is to be used. The review should include a job analysis except as provided in § 12b(3) below with respect to criterion related validity. Any method of job analysis may be used if it provides the information required for the specific validation strategy used.

b. *Criterion-related validity.* (1) Users choosing to validate a selection procedure by a criterion-related validity strategy should determine whether it is technically feasible (as defined in Part IV) to conduct such a study in the particular employment context. The determination of the number of persons necessary to permit the conduct of a meaningful criterion-related study should be made by the user on the basis of all relevant information concerning the selection procedure, the potential sample and the employment situation. These guidelines do not require a user to hire or promote persons for the purpose of making it possible to conduct a criterion-related study; and do not require such a study on a sample of less than thirty (30) persons.

(2) There should be a review of job information to determine measures of work behaviors or performance that are relevant to the job in question. These measures or criteria are relevant to the extent that they represent critical or important job duties, work behaviors or work outcomes as developed from the review of job information. The possibility of bias should be considered both in selection of the measures and their application. In view of the possibility of bias in subjective evaluations, supervisory rating techniques should be carefully developed. All criteria need to be examined for freedom from factors which would unfairly alter scores of members of any group. The relevance of criteria and their freedeom from bias are of particular concern when there are significant differences in measures of job performance for different groups.

(3) Proper safeguards should be taken to insure that scores on selection procedures do not enter into any judgment of employee adequacy that are to be used as criterion measures. Criteria may consist of measures other than work proficiency including, but not limited to length of service, regularity of attendance, training time or properly measured success in job relevant training. Measures of training success based upon pencil and paper tests will be closely reviewed for job relevance. Whatever criteria are used should represent important or critical work behaviors or work outcomes. Job behaviors including but not limited to production rate, error rate, tardiness, absenteeism and turnover, may be used as criteria without a full job analysis if the user can show the importance of the criterion to the particular employment context. A standardized rating of overall work performance may be utilized where a study of the job shows that it is an appropriate criterion.

(4) The sample subjects should insofar as feasible be representative of the candidates normally available in the relevant labor market for the job or jobs in question, and should insofar as feasible include the racial, ethnic and sex groups normally available in the relevant job market. Where samples are combined or compared, attention should be given to see that such samples are comparable in terms of the actual job they perform, the length of time on the job where time on the job is likely to affect performance and other relevant factors likely to affect validity differences; or that these factors are included in the design of the study and their effects identified.

(5) The degree of relationship between selection procedure scores and criterion measures should be examined and computed, using professionally acceptable statistical procedures. Generally, a selection procedure is considered related to the criterion, for the purposes of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the .05 level of significance, which means that it is sufficiently high

as to have a probability of no more than one (1) in twenty (20) to have occurred by chance. Absence of a statistically significant relationship between a selection procedure and job performance should not necessarily discourage other investigations of the validity of that selection procedure.

Users should evaluate each selection procedure to assure that it is appropriate for operational use. Generally, if other factors remain the same, the greater the magnitude of the relationship (e.g., correlation coefficient) between performance on a selection procedure and one or more criteria of performance on the job, and the greater the importance or number of aspects of job performance covered by the criteria, the more likely it is that the procedure will be appropriate for use. Reliance upon a selection procedure which is significantly related to a criterion measure, but which is based upon a study involving a large number of subjects and has a low correlation coefficient will be subject to close review if it has a large adverse impact. Sole reliance upon a single selection instrument which is related to only one of many job duties or aspects of job performance will also be subject to close review. The appropriateness of a selection procedure is best evaluated in each particular situation and there are no minimum correlation coefficients applicable to all employment situations. In determining whether a selection procedure is appropriate for operational use the following considerations should also be taken into account: The degree of adverse impact of the procedure, the availability of other selection procedures of greater or substantially equal validity; and the need of an employer, required by law or regulation to follow merit principles, to have an objective system of selection.

(6) Users should avoid reliance upon techniques which tend to overestimate validity findings as a result of capitalization on chance unless an appropriate safeguard is taken. Reliance upon a few selection procedures or criteria of successful job performance, when many selection procedures or criteria of performance have been studied, or the use of optimal statistical

weights for selection procedures computed in one sample, are techniques which tend to inflate validity estimates as a result of chance. Use of a large sample is one safeguard; cross-validation is another.

(7) *Fairness of the selection procedure.* i. When members of one racial, ethnic, or sex group characteristically obtain lower scores on a selection procedure than members of another group, and the differences are not reflected in differences in measures of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores.

ii. Where a selection procedure results in an adverse impact on a racial, ethnic or sex group identified in accordance with the classifications set forth in § 4 above and that group is a significant factor in the relevant labor market, the user generally should investigate the possible existence of unfairness for that group if it is technically feasible to do so.

The greater the severity of the adverse impact on a group, the greater the need to investigate the possible existence of unfairness. Where the weight of evidence from other studies shows that the selection procedure is a fair predictor for the group in question and for the same or similar jobs, such evidence may be relied on in connection with the selection procedure at issue and may be combined with data from the present study; however, where the severity of adverse impact on a group is significantly greater than in the other studies referred to, a user may not rely on such other studies.

iii. Users conducting a study of fairness should review the APA Standards regarding investigation of possible bias in testing. An investigation of fairness of a selection procedure depends on both evidence of validity and the manner in which the selection procedure is to be used in a particular employment context. Fairness of a selection procedure cannot necessarily be specified in advance without investigating these factors. In-

vestigation of fairness of a selection procedure in samples where the range of scores on selection procedures or criterion measures is severely restricted for any subgroup sample (as compared to other subgroup samples) may produce misleading evidence of unfairness. That factor should accordingly be taken into account in conducting such studies and before reliance is placed on the results.

iv. If unfairness is demonstrated through a showing that members of a particular group perform better or poorer on the job than their scores on the selection procedure would indicate through comparison with how members of other groups perform, the user may either revise or replace the selection instrument in accordance with these guidelines, or may continue to use the selection instrument operationally with appropriate revisions in its use to assure compatibility between the probability of successful job performance and the probability of being selected.

v. In addition to the general conditions needed for technical feasibility for the conduct of a criterion-related study (see § 14(j), below) an investigation of fairness requires the following:

(1) A sufficient number of persons in each group for findings of statistical significance. These guidelines do not require a user to hire or promote persons on the basis of group classifications for the purpose of making it possible to conduct a study of fairness; and do not require a user to conduct a study of fairness on a sample of less than thirty (30) persons for each group involved in the study.

(2) The samples for each group should be comparable in terms of the actual job they perform, length of time on the job where time on the job is likely to affect performance, and other relevant factors likely to affect validity differences; or such factors should be included in the design of the study and their effects identified.

vi. If a study of fairness should otherwise be performed, but is not technically feasible, the use of a selection procedure which has otherwise met the validity standards of these guidelines will be considered in accord with these guidelines, unless the technical infeasibility resulted from discriminatory employment practices which are demonstrated by facts other than past failure to conform with requirements for validation of selection procedures. However, when it becomes technically feasible for the user to perform a study of fairness and such a study is otherwise called for, the user should conduct the study of fairness.

c. *Content validity.* (1) There should be a definition of a performance domain or the performance domains with respect to the job in question. Performance domains may be defined through job analysis, analysis of the work behaviors or activities, or by the pooled judgments of persons having knowledge of the job. Performance domains should be defined on the basis of competent information about job tasks and responsibilities. Performance domains include critical or important work behaviors, work products, work activities, job duties, or the knowledges, skills or abilities shown to be necessary for performance of the duties, behaviors, activities or the production of work. Where a performance domain has been defined as a knowledge, skill or ability, that knowledge, skill or ability must be used in job behavior. A selection procedure based on inferences about psychological processes cannot be supported by content validity alone. Thus content validity by itself is not an appropriate validation strategy for intelligence, aptitude, personality or interest tests. Content validity is also not an appropriate strategy when the selection procedure involves knowledges, skills or abilities which an employee will be expected to learn on the job.

(2) If a higher score on a content valid selection procedure can be expected to result in better job performance the results may be used to rank persons who score above minimum levels. Where a selection procedure supported solely by content va-

lidity is used to rank job candidates, the performance domain should include those aspects of performance which differentiate among levels of job performance.

(3) A selection procedure which is a representative sample of a performance domain of the job as defined in accordance with subsection (1) above, is a content valid procedure for that domain. Where the domain or domains measured are critical to the job, or constitute a substantial proportion of the job, the selection procedure will be considered to be content valid for the job. The reliability of selection procedures justified on the basis of content validity should be a matter of concern to the user. Whenever it is feasible to do so, appropriate statistical estimates should be made of the reliability of the selection procedures.

(4) A demonstration of the relationship between the content of the selection procedure and the performance domain of the job is critical to content validity. Content validity may be shown if the knowledges, skills or abilities demonstrated in and measured by the selection procedure are substantially the same as the knowledges, skills or abilities shown to be necessary for job performance. The closer the content of the selection procedure is to actual work samples, behaviors or activities, the stronger is the basis for showing content validity. The need for careful documentation of the relationship between the performance domain of the selection procedure and that of the job increases as the content of the selection procedure less resembles that of the job performance domain.

(5) A requirement for specific prior training or for work experience based on content validity, including a specification of level or amount of training or experience, should be justified on the basis of the relationship between the content of the training or experience and the performance domain of the job for which the training or experience is to be required.

(6) If a selection procedure is supported solely on the basis of content validity, it may be used operationally if it represents a

critical performance domain or a substantial proportion of the performance domains of the job.

d. *Construct validity.* Construct validity is a more complex strategy than either criterion-related or content validity. Accordingly, users choosing to validate a selection procedure by use this strategy should be careful to follow profesionally accepted standards, such as those contained in the APA standards and the standard text books and journals.

(1) There should be a job analysis. This job analysis should result in a determination of the constructs that underlie successful performance of the important or critical duties of the job.

(2) A selection procedure should be selected or developed which measures the construct(s) identified in accord with subparagraph (1) above.

(3) A selection procedure may be used operationally if the standards of subparagraphs (1) and (2) are met and there is sufficient empirical research evidence showing that the procedure is validity related to performance of critical job duties. Normally, sufficient empirical research evidence would take the form of one or more criterion related validity studies meeting the requirements of § 12b. See also second sentence of § 12.

(4) Where a selection procedure satisfies the standards of subsections (1), (2) and (3) above, it may be used operationally for other jobs which are shown by an appropriate job analysis to include the same construct(s) as an essential element in job performance.

PART III—DOCUMENTATION OF VALIDITY EVIDENCE

§ 13a. For each selection procedure having an adverse impact (as set forth in § 4) the user should maintain and have available

the data on which the adverse impact determination was made and one of the following types of documentation evidence:

(1) Documentation evidence showing criterion related validity of the selection procedure (see § 13b. infra).

(2) Documentation evidence showing content validity of the selection procedure (see § 13c. infra).

(3) Documentation evidence showing construct validity of the selection procedure (see § 13d. infra).

(4) Documentation evidence from other studies showing validity of the selection procedure in the user's facility (see § 13e. infra).

(5) Documentation evidence showing what steps were taken to reduce or eliminate adverse impact, why validation is not feasible or not appropriate and why continued use of the procedure is consistent with Federal law.

This evidence should be compiled in a reasonably complete and organized manner to permit direct evaluation of the validity of the selection procedure. Previously written employer or consultant reports of validity are acceptable if they are complete in regard to the following documentation requirements, or if they satisfied requirements of guidelines which were in effect when the study was completed. If they are not complete, the required additional documentation should be appended. If necessary information is not available the report of the validity study may still be used as documentation, but its adequacy will be evaluated in terms of compliance with the requirements of these guidelines.

In the event that evidence of validity is reviewed by an enforcement agency, the reports completed after the effective date of these guidelines are expected to use one of the formats set forth below.

Evidence denoted by use of the word "(Essential)" is considered critical and reports not containing such information will

be considered incomplete. Evidence not so denoted is desirable, but its absence will not be a basis for considering a report incomplete.

b. *Criterion-related validity.* Reports of criterion-related validity of selection procedures are to contain the following information:

(1) *User(s), and location(s) and date(s) of study.* Dates of administration of selection procedures and collection of criterion data and, where appropriate, the time between collection of data on selection procedures and criterion measures should be shown (Essential). If the study was conducted at several locations, the address of each location, including city and state, should be shown.

(2) *Problem and setting.* An explicit definition of the purpose(s) of the study and the circumstances in which the study was conducted should be provided. A description of existing selection procedures and cut-off scores, if any, should be provided.

(3) *Review of job information or job analysis.* Where a review of job information results in criteria which are measures other than work proficiency (see 12b(3)), the basis for the selection of these criteria should be reported (Essential). Where a job analysis is required, the report should include either: (a) The important duties performed on the job and the basis on which such duties were determined to be important, such as the proportion of time spent on the respective duties, their level of difficulty, their frequency of performance, the consequences of error, or other appropriate factors; or (b) the knowledge, skills, abilities and/or other worker characteristics and bases on which they were determined to be important for job performance (Essential). Published descriptions from industry sources or volume I of the Dictionary of Occupational Titles Third Edition, United States Government Printing Office, 1965, are satisfactory if

they adequately and completely describe the job. If appropriate, a brief supplement to the published description should be provided.

If two or more jobs are grouped for a validity study, a justification for this grouping, as well as a description of each of the jobs, should be provided (Essential).

(4) *Job titles and codes.* It is desirable to provide the user's job title(s) for the job(s) in question and the corresponding job title(s) and code(s) from United States Employment Service Dictionary of Occupational Titles Volumes I and II. Where standard titles and codes do not exist, a notation to that effect should be made.

(5) *Criteria.* A full description of all criteria on which data were collected including a rationale for selection of the final criteria, and means by which they were observed, recorded, evaluated and quantified, should be provided (Essential). If rating techniques are used as criterion measures the appraisal form(s) and instructions to the rater(s) should be included as part of the validation evidence (Essential).

(6) *Sample.* A description of how the research sample was selected should be included (Essential). The racial, ethnic and sex composition of the sample should be described, including the size of each subgroup (Essential). Racial and ethnic classifications should be those set forth in § 4a above. A description of how the research sample compares with the racial, ethnic and sex composition of the relevant labor market is also desirable. Where data are available, the racial, ethnic and sex composition of current applicants should also be described. Descriptions of educational levels, length of service, and age are also desirable.

(7) *Selection procedure.* Any measure, combination of measures, or procedures used as a basis for employment decisions should be completely and explicitly described or attached

(Essential). If commercially available selection procedures are used, they should be described by title, form, and publisher (Essential). Reports of reliability estimates and how they were established are desirable. A rationale for choosing the selection procedures investigated in the study should be included.

(8) *Techniques and results.* Methods used in analyzing data should be described (Essential). Measures of central tendency (e.g., means) and measures of dispersion (e.g., standard deviations and ranges) for all selection procedures and all criteria should be reported for all relevant racial, ethnic and sex subgroups (Essential). Statistical results should be organized and presented in tabular or graphical form, by racial, ethnic and/or sex subgroups (Essential). All selection procedure-criterion relationships investigated should be reported, including their magnitudes and directions (Essential). Statements regarding the statistical significance of results should be made (Essential).

Any statistical adjustments, such as for less than perfect reliability or for restriction of score range in the selection procedure or criterion, or both, should be described; and uncorrected correlation coefficients should also be shown (Essential). Where the statistical technique used categorizes continuous data, such as bi-serial correlation and the phi coefficient, the categories and the bases on which they were determined should be described (Essential). Studies of test fairness should be included where called for by the requirements of § 12b(7) (Essential). These studies should include the rationale by which a selection procedure was determined to be fair to the group(s) in question. Where test fairness has been demonstrated on the basis of other studies, a bibliography of the relevant studies should be included (Essential). If the bibliography includes unpublished studies, copies of these studies, or adequate abstracts or summaries, should be attached (Essential). Where revisions have been made in a selection procedure to assure compatability between successful job performance and the probability of be-

ing selected, the studies underlying such revisions should be included (Essential).

(9) *Uses and applications.* A description of the way in which each selection procedure is to be used (e.g., as a screening device with a cut-off score or combined with other procedures in a battery) and application of the procedure (e.g., selection, transfer, promotion) should be provided (Essential). If weights are assigned to different parts of the selection procedure, these weights and the validity of the weighted composite should be reported (Essential).

(10) *Cut-off scores.* Where cut-off scores are to be used, both the cut-off scores and the way in which they were determined should be described (Essential).

(11) *Source data.* Each user should maintain records showing all pertinent information about individual sample members in studies involving the validation of selection procedures. These records (exclusive of names and social security number) should be made available upon request of a compliance agency. These data should include selection procedure scores, criterion scores, age, sex, minority group status, and experience on the specific job on which the validation study was conducted and may also include such things as education, training, and prior job experience. If the user chooses to include, along with a report on validation, a worksheet showing the pertinent information about the individual sample members, specific identifying information such as name and social security number should not be shown. Inclusion of the worksheet with the validity report is encouraged in order to avoid delays.

(12) *Contact person.* It is desirable for the user to set forth the name, mailing address, and telephone number of the individual who may be contacted for further information about the validity study.

c. *Content validity.* Reports of content validity of selection procedures are to contain the following information:

(1) *Definition of performance domain.* A full description should be provided for the basis on which a performance domain is defined (Essential). A complete and comprehensive definition of the performance domain should also be provided (Essential). The domain should be defined on the basis of competent information about job tasks and responsibilities (Essential). Where the performance domain is defined in terms of knowledges, skills, or abilities, there should be an operational definition of each knowledge, skill or ability and a complete description of its relationship to job duties, behaviors, activities, or work products (Essential).

(2) *Job title and code.* It is desirable to provide the user's job title(s) and the corresponding job title(s) and code(s) from the United States Employment Service Dictionary of Occupational Titles Volumes I and II. Where standard titles and codes do not exist, a notation to that effect should be made.

(3) *Selection procedures.* Selection procedures including those constructed by or for the user, specific training requirements, composites of selection procedures, and any other procedure for which content validity is asserted should be completely and explicitly described or attached (Essential). If commercially available selection procedures are used, they should be described by title, form, and publisher (Essential). Where the performance domain is defined in terms of knowledges, skills or abilities, evidence that the selection procedure measures those knowledges, skills and abilities should be provided (Essential).

(4) *Techniques and results.* The method by which the correspondence between the content of the selection procedure and the job performance domain(s) was established and the relative emphasis given to various aspects of the content of the selection procedure as derived from the performance domain(s) should be described (Essential). If any steps were taken to reduce adverse racial, ethnic, or sex impact in the content of the procedure or in its administration, these steps should be described.

Establishment of time limits, if any, and how these limits are related to the speed with which duties must be performed on the job, should be explained, the adequacy of the sample coverage of the performance domain should be described as precisely as possible. Measures of central tendency (e.g., means) and measures of dispersion (e.g., standard deviations) should be reported for all selection procedures as appropriate. Such reports should be made for all relevant racial, ethnic, and sex subgroups, at least on a statistically reliable sample basis.

(5) *Uses and applications.* A description of the way in which each selection procedure is to be used (e.g., as a screening device with a cut-off score or combined with other procedures in a battery) and the application of the procedure (e.g., selection, transfer, promotion) should be provided (Essential).

(6) *Cut-off scores.* The rationale for minimum scores, if used, should be provided (Essential). If the selection procedure is used to rank individuals above minimum levels, or if preference is given to individuals who score significantly above the minimum levels, a rationale for this procedure should be provided (Essential).

(7) *Contact person.* It is desirable for the employer to set forth the name, mailing address and telephone number of the individual who may be contacted for further information about the validation study.

d. *Construct validity.* Reports of construct validity of selection procedures are to contain the following information:

(1) *Construct definition.* A clear definition of the construct should be provided, explained in terms of empirically observable behavior, including levels of construct performance relevant to the job(s) for which the selection procedure is to be used (Essential).

(2) *Job analysis.* The job analysis should show how the construct underlying successful job performance of important or

critical duties were determined (Essential). The job analysis should provide evidence of the linkage between the construct and the important duties of the job and how this linkage was determined (Essential).

(3) *Job titles and codes.* It is desirable to provide the selection procedure user's job title(s) for the job(s) in question and the corresponding job title(s) and code(s) from the United States Employent Service Dictionary of Occupational Titles, Volume I and II. Where standard titles and codes do not exist, a notation to that effect should be made.

(4) *Selection procedure.* The selection procedure used as a measure of the construct should be completely and explicitly described or attached (Essential). If commercially available selection procedures are used, they should be identified by title, form and publisher (Essential). The evidence demonstrating that the selection procedure is in fact a proper measure of the construct should be included (Essential). Reports of reliability estimates and how they were established are desirable.

(5) *Anchoring.* The empirical evidence showing that performance on the selection procedure is validly related to performance of critical job duties should be included (Essential).

(6) *Uses and applications.* A description of the way in which each selection procedure is to be used (e.g., as a screening device with a cut-off score or combined with other procedures in a battery) and application of the procedure (e.g., selection, transfer, promotion) should be provided (Essential). If weights are assigned to different parts of the selection procedure, these weights (and the validity of the weighted composite) should be reported (Essential).

(7) *Cut-off scores.* Where cut-off scores are to be used, both the cut-off scores and the way in which they were determined should be described (Essential).

(8) *Source data.* Each user should maintain records showing all pertinent information about individual sample members in studies involving the validation of selection procedures. These records (exclusive of names and social security number) should be made available upon request of a compliance agency. These data should include selection procedure scores, criterion scores, age, sex, minority group status, and experience on the specific job on which the validation study was conducted and may also include such things as education, training, and prior job experience. If the user chooses to include, along with a report on validation, a worksheet showing the pertinent information about the individual sample members, specific identifying information such as name and social security number should not be shown. Inclusion of the worksheet with the validity report is encouraged in order to avoid delays.

(9) *Contact person.* It is desirable for the user to set forth the name, mailing address, and telephone number of the individual who may be contacted for further information about the validity study.

e. *Evidence of validity from other studies.* When validity of a selection procedure is supported by studies not done by the user, the evidence from the original study or studies should be compiled in a manner similar to that required in the appropriate section of this § 13 above. In addition, the following evidence should be supplied:

(1) *Evidence from criterion-related validity studies.* (i) *Job Information.*— A description of the important duties of the user's job and the basis on which the duties were determined to be important should be provided (Essential). A full description of the basis for determining that these important job duties are sufficiently similar to the duties of the job in the original study (or studies) to warrant use of the selection procedure in the new situation should be provided (Essential).

(ii) *Relevance of criteria.*—A full description of the basis on which the criteria used in the original studies are determined to be relevant for the user should be provided (Essential).

(iii) *Other variables.*—The similarity of important applicant pool/sample characteristics reported in the original studies to those of the user should be described (Essential). A description of the comparison between the race and sex composition of the user's relevant labor market and the sample in the original validity studies should be provided (Essential).

(iv) *Use of the selection procedure.*—A full description should be provided showing that the use to be made of the selection procedure is consistent with the findings of the original validity studies (Essential).

(v) *Bibliography.*—A bibliography of reports of validity of the selection procedure for the job or jobs in question should be provided (Essential). Where any of the studies included an investigation of test fairness, the results of this investigation should be provided (Essential). Copies of reports published in journals that are not commonly available should be described in detail or attached (Essential). Where a user is relying upon unpublished studies, a reasonable effort should be made to obtain these studies. If these unpublished studies are the sole source of validity evidence they should be described in detail or attached (Essential). If these studies are not available, the name and address of the source, an adequate abstract or summary of the validity study and data, and a contact person in the source organization should be provided (Essential).

(2) *Evidence from content validity studies.* (i) *Similarity of Performance Domains.*—A full description should be provided of the similarity between the performance domain in the user's job and the performance domain measured by a selection procedure developed and shown to be content valid by another user (Essential). The basis for determining this similarity should be explicitly described (Essential).

(3) *Evidence from construct validity studies.* (i) *Uniformity of Construct.*—A full description should be provided of the basis for determining that the construct identified as underlying successful job performance by the user's job analysis is the same as the construct measured by the selection procedure (Essential).

PART IV—DEFINITIONS

§ 14 The following defintions shall apply throughout these guidelines:

(a) Ability: The present observable competence to perform a function.

(b) Adverse Impact: Defined in § 4 of these guidelines.

(c) Employer: Any employer subject to the provisions of the Civil Rights Act of 1964, as amended, including state or local governments and any Federal agency subject to the provisions of Sec. 717 of the Civil Rights Act of 1964, as amended, and any Federal contractor or subcontractor or federally assisted construction contractor or subcontractor covered by Executive Order 11246, as amended.

(d) Employment agency: Any employment agency subject to the provisions of the Civil Rights Act of 1964, as amended.

(e) Labor organization: Any labor organization subject to the provisions of the Civil Rights Act of 1964, as amended, and any committee controlling apprenticeship or other training.

(f) Enforcement agency: Any agency of the executive branch of the Federal Government which adopts these guidelines for purposes of the enforcement of the equal employment opportunity laws or which has responsibility for securing compliance with them.

(g) Labor organization: Any labor organization subject to the provisions of the Civil Rights Act of 1964, as amended, and any committee controlling apprenticeship or other training.

(h) Racial, sex or ethnic group: Any group of persons identifiable on the grounds of race, color, religion, sex or national origin.

(i) Selection procedure: Any measure, combination of measures, or procedure, other than a bona fide seniority system, used as a basis for any employment decision. Selection procedures include the full range of assessment techniques from traditional paper and pencil tests, performance tests, training programs or probationany periods and physical, educational and work experience requirements through informal or casual interviews and unscored application forms.

(j) Selection rate: The proportion of applicants or candidates who are hired, promoted or otherwise selected.

(k) Technical feasibility: The existence of conditions permitting the conduct of meaningful criterion related validity studies. These conditions include: (a) An adequate sample of persons available for the study to achieve findings of statistical significance; (b) having or being able to obtain a sufficient range of scores on the selection procedure and job performance measures to produce validity results which can be expected to be representative of the results if the ranges normally expected were utilized; and (c) having or being able to devise unbiased, reliable and relevant measures of job performance or other criteria of employee adequacy. See § 12b(1). With respect to investigation of possible unfairness, the same considerations are applicable to each group for which the study is made. See §12b(7).

(1) Unfairness of selection procedure (differential prediction): A condition in which members of one racial, ethnic, or sex group characteristically obtain lower scores on a selection procedure than members of another group, and the differences are not reflected in differences in measures of job performance. See § 12b(7).

(m) User: Any employer, labor organization, employment agency, or licensing or certification board, to the extent it may be covered by Federal equal employment opportunity law which uses a selection procedure as a basis for any employment decision. Whenever an employer, labor organization, or employment agency is required by law to restrict recruitment for any occupation to those applicants who have met licensing or certification requirements, the licensing or certifying authority to the extent it may be covered by Federal equal employment opportunity law will be considered the user with respect to those licensing or certification requirements. Whenever a state employment agency or service does no more than administer or monitor a procedure as permitted by Department of Labor regulations, and does so without making referrals or taking any other action on the basis of the results, the state employment agency will not be deemed to be a user.

APPENDIX H

Equal Employment Opportunity Commission Guidelines On Employment Selection Procedures 29 CF.R §1607 (1970)

Table of Contents

1607.1 Statement of Purpose
1607.2 "Test" Defined
1607.3 Discrimination Defined
1607.4 Evidence of Validity
1607.5 Minimum Standards for Validation
1607.6 Presentation of Validity Evidence
1607.7 Use of Other Validity Studies
1607.8 Assumption of Validity
1607.9 Continued Use of Tests
1607.10 Employment Agencies and Employment Services
1607.11 Disparate Treatment
1607.12 Retesting
1607.13 Other Selection Techniques
1607.14 Affirmative Action

Section 1607.1 Statement of Purpose.—(a) The guidelines in this part are based on the belief that properly validated and standardized employee selection procedures can significantly contribute to the implementation of nondiscriminatory personnel policies, as required by Title VII. It is also recognized that professionally developed tests, when used in conjunction with other tools of personnel assessment and complemented by sound programs of job design, may significantly aid in the development and maintenance of an efficient work force and, indeed, aid in the utilization and conservation of human resources generally.

(b) An examination of charges of discrimination filed with the Commission and an evaluation of the results of the Commis-

sion's compliance activities has revealed a decided increase in total test usage and a marked increase in doubtful testing practices which, based on our experience, tend to have discriminatory effects. In many cases, persons have come to rely almost exclusively on tests as the basis for making the decision to hire, transfer, promote, grant membership, train, refer or retain, with the result that candidates are selected or rejected on the basis of a single test score. Where tests are so used, minority candidates frequently experience disproportionately high rates of rejection by failing to attain score levels that have been established as minimum standards for qualification.

It has also become clear that in many instances persons are using tests as the basis for employment decisions without evidence that they are valid predictors of employee job performance. Where evidence in support of presumed relationships between test performance and job behavior is lacking, the possibility of discrimination in the application of test results must be recognized. A test lacking demonstrated validity (i.e., having no known significant relationship to job behavior) and yielding lower scores for classes protected by Title VII may result in the rejection of many who have necessary qualifications for successful work performance.

(c) The guidelines in this part are designed to serve as a workable set of standards for employers, unions and employment agencies in determining whether their selection procedures conform with the obligations contained in Title VII of the Civil Rights Act of 1964. Section 703 of Title VII places an affirmative obligation upon employers, labor unions, and employment agencies, as defined in section 701 of the Act, not to discriminate because of race, color, religion, sex, or national origin. Subsection (h) of section 703 allows such persons "*** to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin."

Sec. 1607.2. "Test" Defined.—For the purpose of the guidelines in this part, the term "test" is defined as any paper-and-pencil or performance measure used as a basis for any employment decision. The guidelines in this part apply, for example, to ability tests which are designed to measure eligibility for hire, transfer, promotion, membership, training, referral or retention. This definition includes, but is not restricted to, measures of general intelligence, mental ability and learning ability; specific intellectual abilities; mechanical, clerical and other aptitudes; dexterity and coordination; knowledge and proficiency; occupational and other interests; and attitudes, personality or temperament. The term "test" includes all formal, scored, quantified or standardized techniques of assessing job suitability including, in addition to the above, specific qualifying or disqualifying personal history or background requirements, specific educational or work history requirements, scored interviews, biographical information blanks, interviewers' rating scales, scored application forms, etc.

Sec. 1607.3. Discrimination Defined.—The use of any test which adversely affects hiring, promotion, transfer or any other employment or membership opportunity of classes protected by Title VII constitutes discrimination unless: (a) the test has been validated and evidences a high degree of utility as hereinafter described, and (b) the person giving or acting upon the results of the particular test can demonstrate that alternative suitable hiring, transfer or promotion procedures are unavailable for his use.

Sec. 1607.4. Evidence of Validity.—(a) Each person using tests to select from among candidates for a position or for membership shall have available for inspection evidence that the tests are being used in a manner which does not violate § 1607.3. Such evidence shall be examined for indications of possible discrimination, such as instances of higher rejection rates for minority candidates than nonminority candidates. Furthermore, where technically feasible, a test should be validated for each

minority group with which it is used; that is, any differential rejection rates that may exist, based on a test, must be relevant to performance on the jobs in question.

(b) The term "technically feasible" as used in these guidelines means having or obtaining a sufficient number of minority individuals to achieve findings of statistical and practical significance, the opportunity to obtain unbiased job performance criteria, etc. It is the responsibility of the person claiming absence of technical feasibility to positively demonstrate evidence of this absence.

(c) Evidence of a test's validity should consist of empirical data demonstrating that the test is predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.

(1) If job progression structures and seniority provisions are so established that new employees will probably, within a reasonable period of time and in a great majority of cases, progress to a higher level, it may be considered that candidates are being evaluated for jobs at that higher level. However, where job progression is not so nearly automatic, or the time span is such that higher level jobs or employees' potential may be expected to change in significant ways, it shall be considered that candidates are being evaluated for a job at or near the entry level. This point is made to underscore the principle that attainment of or performance at a higher level job is a relevant criterion in validating employment tests only when there is a high probability that persons employed will in fact attain that higher level job within a reasonable period of time.

(2) Where a test is to be used in different units of a multiunit organization and no significant differences exist between units, jobs, and applicant populations, evidence obtained in one unit may suffice for the others. Similarly, where the validation process requires the collection of data throughout a multiunit

organization, evidence of validity specific to each unit may not be required. There may be also instances where evidence of validity is appropriately obtained from more than one company in the same industry. Both in this instance and in the use of data collected throughout a multiunit organization, evidence of validity specific to each unit may not be required: *Provided,* That no significant differences exist between units, jobs, and applicant populations.

Sec. 1607.5. Minimum Standards for Validation.—(a) For the purpose of satisfying the requirements of this part, empirical evidence in support of a test's validity must be based on studies employing generally accepted procedures for determining criterion-related validity, such as those described in "Standards for Educational and Psychological Tests and Manuals" published by American Psychological Association, 1200 17th Street NW., Washington, D.C. 20036. Evidence of content or construct validity, as defined in that publication, may also be appropriate where criterion-related validity is not feasible. However, evidence for content or construct validity should be accompanied by sufficient information from job analyses to demonstrate the relevance of the content (in the case of job knowledge or proficiency tests) or the construct (in the case of trait measures). Evidence of content validity alone may be acceptable for well-developed tests that consist of suitable samples of the essential knowledge, skills or behaviors composing the job in question. The types of knowledge, skills or behaviors contemplated here do not include those which can be acquired in a brief orientation to the job.

(b) Although any appropriate validation strategy may be used to develop such empirical evidence, the following minimum standards, as applicable, must be met in the research approach and in the presentation of results which constitute evidence of validity.

(1) Where a validity study is conducted in which tests are administered to applicants, with criterion data collected later, the

sample of subjects must be represenative of the normal or typical candidate group for the job or jobs in question. This further assumes that the applicant sample is representative of the minority population available for the job or jobs in question in the local labor market. Where a validity study is conducted in which tests are administered to present employees, the sample must be representative of the minority groups currently included in the applicant population. If it is not technically feasible to include minority employees in validation studies conducted on the present work force, the conduct of a validation study without minority candidates does not relieve any person of his subsequent obligation for validation when inclusion of minority candidates becomes technically feasible.

(2) Tests must be administered and scored under controlled and standardized conditions, with proper safeguards to protect the security of test scores and to insure that scores do not enter into any judgments of employee adequacy that are to be used as criterion measures. Copies of tests and test manuals, including instructions for administration, scoring, and interpretation of test results, that are privately developed and/or are not available through normal commercial channels must be included as a part of the validation evidence.

(3) The work behaviors or other criteria of employee adequacy which the test is intended to predict or identify must be fully described; and, additionally, in the case of rating techniques, the appraisal form(s) and instructions to the rater(s) must be included as a part of the validation evidence. Such criteria may include measures other than actual work proficiency, such as training time, supervisory ratings, regularity of attendance and tenure. Whatever criteria are used they must represent major or critical work behaviors as revealed by careful job analyses.

(4) In view of the possibility of bias inherent in subjective evaluations, supervisory rating techniques should be carefully

developed, and the ratings should be closely examined for evidence of bias. In addition, minorities might obtain unfairly low performance criterion scores for reasons other than supervisors' prejudice, as, when. as new employees, they have had less opportunity to learn job skills. The general point is that all criteria need to be examined to insure freedom from factors which would unfairly depress the scores of minority groups.

(5) Differential validity. Data must be generated and results separately reported for minority and nonminority groups wherever technically feasible. Where a minority group is sufficiently large to constitute an identifiable factor in the local labor market, but validation data have not been developed and presented separately for that group, evidence of satisfactory validity based on other groups will be regarded as only provisional compliance with these guidelines pending separate validation of the test for the minority group in question. (See § 1607.9). A test which is differentially valid may be used in groups for which it is valid but not for those in which it is not valid. In this regard where a test is valid for two groups but one group characteristically obtains higher test scores than the other without a corresponding difference in job performance, cutoff scores must be set so as to predict the same probability of job success in both groups.

(c) In assessing the utility of a test the following considerations will be applicable:

(1) The relationship between the test and at least one relevant criterion must be statistically significant. This ordinarily means that the relationship should be sufficiently high as to have a probability of no more than 1 to 20 to have occurred by chance. However, the use of a single test as the sole selection device will be scrutinized closely when that test is valid against only one component of job performance.

(2) In addition to statistical significance, the relationship between the test and criterion should have practical significance.

The magnitude of the relationship needed for practical significance or usefulness is affected by several factors, including:

(i) The larger the proportion of applicants who are hired for or placed on the job, the higher the relationship needs to be in order to be practically useful. Conversely, a relatively low relationship may prove useful when proportionately few job vacancies are available;

(ii) The larger the proportion of applicants who become satisfactory employees when not selected on the basis of the test, the higher the relationship needs to be between the test and a criterion of job success for the test to be practically useful. Conversely, a relatively low relationship may prove useful when proportionately few applicants turn out to be satisfactory;

(iii) The smaller the economic and human risks involved in hiring an unqualified applicant relative to the risks entailed in rejecting a qualified applicant, the greater the relationship needs to be in order to be practically useful. Conversely, a relatively low relationship may prove useful when the former risks are relatively high.

Sec. 1607.6. Presentation of Validity Evidence.—The presentation of the results of a validation study must include graphical and statistical representations of the relationships between the test and the criteria, permitting judgments of the test's utility in making predictions of future work behavior. (See §1607.5(c) concerning assessing utility of a test.) Average scores for all tests and criteria must be reported for all relevant subgroups, including minority and nonminority groups where differential validation is required. Whenever statistical adjustments are made in validity results for less than perfect reliability or for restriction of score range in the test or the criterion, or both, the supporting evidence from the validation study must be presented in detail. Furthermore, for each test that is to be established or continued as an operational employee selection instrument, as a result of the validation study, the minimum ac-

ceptable cutoff (passing) score on the test must be reported. It is expected that each operational cutoff score will be reasonable and consistent with normal expectations of proficiency within the work force or group on which the study was conducted.

Sec. 1607.7. Use of Other Validity Studies.—In cases where the validity of a test cannot be determined pursuant to §1607.4 and §1607.5 (e. g., the number of subjects is less than that required for a technically adequate validation study, or an appropriate criterion measure cannot be developed), evidence from validity studies conducted in other organizations, such as that reported in test manuals and professional literature, may be considered acceptable when: (a) The studies pertain to jobs which are comparable (i. e., have basically the same task elements), and (b) there are no major differences in contextual variables or sample composition which are likely to significantly affect validity. Any person citing evidence from other validity studies as evidence of test validity for his own jobs must substantiate in detail job comparability and must demonstrate the absence of contextual or sample differences cited in paragraphs (a) and (b) of this section.

Sec. 1607.8. Assumption of Validity.—(a) Under no circumstances will the general reputation of a test, its author or its publisher, or casual reports of test utility be accepted in lieu of evidence of validity. Specifically ruled out are: assumptions of validity based on test names or descriptive labels; all forms of promotional literature; data bearing on the frequency of a test's usage; testimonial statements of sellers, users, or consultants; and other nonempirical or anecdotal accounts of testing practices or testing outcomes.

(b) Although professional supervision of testing activities may help greatly to insure technically sound and nondiscriminatory test usage, such involvement alone shall not be regarded as constituting satisfactorily evidence of test validity.

Sec. 1607.9. Continued Use of Tests.—Under certain conditions, a person may be permitted to continue the use of a test which is not at the moment fully supported by the required evidence of validity. If, for example, determination of criterion-related validity in a specific setting is practicable and required but not yet obtained, the use of the test may continue: *Provided:* (a) The person can cite substantial evidence of validity as described in §1607.7(a) and (b); and (b) he has in progress validation procedures which are designed to produce, within a reasonable time, the additional data required. It is expected also that the person may have to alter or suspend test cutoff scores so that score ranges broad enough to permit the identification of criterion-related validity will be obtained.

Sec. 1607.10. Employment Agencies and Employment Services.—(a) An employment service, including private employment agencies, State employment agencies, and the U. S. Training and Employment Service, as defined in section 701(c), shall not make applicant or employee appraisals or referrals based on the results obtained from any psychological test or other selection standard not validated in accordance with these guidelines.

(b) An employment agency or service which is requested by an employer or union to devise a testing program is required to follow the standards for test validation as set forth in these guidelines. An employment service is not relieved of its obligation herein because the test user did not request such validation or has requested the use of some lesser standard than is provided in these guidelines.

(c) Where an employment agency or service is requested only to administer a testing program which has been elsewhere devised, the employment agency or service shall request evidence of validation, as described in the guidelines in this part, before it administers the testing program and/or makes referral pursuant to the test results. The employment agency must furnish on request such evidence of validation. An employment agency or

service will be expected to refuse to administer a test where the employer or union does not supply satisfactory evidence of validation. Reliance by the test user on the reputation of the test, its author, or the name of the test shall not be deemed sufficient evidence of validity (see §1607.8(a)). An employment agency or service may administer a testing program where the evidence of validity comports with the standards provided in §1607.7.

Sec. 1607.11. Disparate Treatment.—The principle of disparate or unequal treatment must be distinguished from the concepts of test validation. A test or other employee selection standard—even though validated against job performance in accordance with the guidelines in this part—cannot be imposed upon any individual or class protected by Title VII where other employees, applicants or members have not been subjected to that standard. Disparate treatment, for example, occurs where members of a minority or sex group have been denied the same employment, promotion, transfer or membership opportunities as have been made available to other employees or applicants. Those employees or applicants who have been denied equal treatment, because of prior discriminatory practices or policies, must at least be afforded the same opportunities as had existed for other employees or applicants during the period of discrimination. Thus, no new test or other employee selection standard can be imposed upon a class of individuals protected by Title VII who, but for prior discrimination, would have been granted the opportunity to qualify under less stringent selection standards previously in force.

Sec. 1607.12. Retesting.—Employers, unions, and employment agencies should provide an opportunity for retesting and reconsideration to earlier "failure" candidates who have availed themselves of more training or experience. In particular, if any applicant or employee during the course of an interview or other employment procedure claims more education or experience, that individual should be retested.

Sec. 1607.13. Other Selection Techniques.—Selection techniques other than tests, as defined in § 1607.2, may be improperly used so as to have the effect of discriminating against minority groups. Such techniques include, but are not restricted to, unscored or casual interviews and unscored application forms. Where there are data suggesting employment discrimination, the person may be called upon to present evidence concerning the validity of his unscored procedures as well as of any tests which may be used, the evidence of validity being of the same types referred to in §§ 1607.4 and 1607.5. Data suggesting the possibility of discrimination exist, for example, when there are differential rates of applicant rejection from various minority and non-minority or sex groups for the same job or group of jobs or when there are disproportionate representations of minority and nonminority or sex groups among present employees in different types of jobs. If the person is unable or unwilling to perform such validation studies, he has the option of adjusting employment procedures so as to eliminate the conditions suggestive of employment discrimination.

Sec. 1607.14. Affirmative Action.—Nothing in these guidelines shall be interpreted as diminishing a person's obligation under both Title VII and Executive Order 11246 as amended by Executive Order 11375 to undertake affirmative action to ensure that applicants or employees are treated without regard to race, color, religion, sex, or national origin. Specifically, the use of tests which have been validated pursuant to these guidelines does not relieve employers, unions or employment agencies of their obligations to take positive action in affording employment and training to members of classes protected by Title VII.

The guidelines in this part are effective upon publication in the *Federal Register* [35 F.R. 12333, August 1, 1970].

Supreme Court, U. S.
FILED
OCT 15 1980
MICHAEL RODAK, JR., CLERK

No. 80-406

IN THE

Supreme Court of the United States

OCTOBER TERM, 1980

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

v.

THE UNITED STATES OF AMERICA *et al.*,
Respondents.

On Petition for Writ of Certiorari to the United States Court of Appeals for the Fifth Circuit

---

BRIEF FOR RESPONDENTS JOHN W. MARTIN *ET AL.* IN OPPOSITION

RICHARD M. SHARP
WILLIAM R. GALEOTA
SHEA & GARDNER
1800 Massachusetts Avenue, N.W.
Washington, D.C. 20036
(202) 828-2000

Counsel of Record for Respondents John W. Martin et al.

Of Counsel:

WILLIAM L. ROBINSON
RICHARD T. SEYMOUR
National Lawyers' Committee
for Civil Rights Under Law
733 Fifteenth Street, N.W.
Suite 520
Washington, D.C. 20005

WILSON - EPES PRINTING CO., INC. - 789-0096 - WASHINGTON, D.C. 20001

QUESTION PRESENTED

Whether the court of appeals correctly affirmed the district court's finding that the Personnel Board of Jefferson County, Alabama, had not met its burden, under Title VII of the Civil Rights Act of 1964, of showing that the employment tests at issue reliably predicted job performance of applicants for jobs of police officer and firefighter.

TABLE OF CONTENTS

	Page
STATEMENT	2
A. Adverse Impact	2
B. Job-Relatedness	3
C. *Washington* v. *Davis*	7
SUMMARY OF ARGUMENT	7
ARGUMENT	8
I. THE DECISION BELOW DOES NOT CONFLICT WITH *WASHINGTON* v. *DAVIS*	8
II. THE BOARD'S ARGUMENT THAT THE DISTRICT COURT ERRED IN BALANCING THE RACIAL IMPACT OF THE POLICE OFFICERS' TEST AGAINST ITS VALIDITY DOES NOT WARRANT REVIEW	13
A. The District Court Did Not "Balance" The Racial Impact Of The Test Against Its Validity	13
B. Title VII Permits A Court To Assess The Racial Impact Of An Employment Practice When Considering Whether It Is Appropriate For Use	15
CONCLUSION	17

PRECEDING PAGE WAS BLANK

TABLE OF AUTHORITIES

Cases: Page

Albemarle Paper Co. v. *Moody*, 422 U.S. 405 (1975) 12
Berenyi v. *Immigration Director*, 385 U.S. 630 (1967) 11
Blake v. *City of Los Angeles*, 595 F.2d 1367 (9th Cir. 1979), *cert. denied*, 48 U.S.L.W. 3698 (1980) 12
Davis v. *Washington*, 348 F. Supp. 15 (D.D.C. 1972) 9
Dozier v. *Chupka*, 395 F. Supp. 836 (S.D. Ohio 1975) 12
Graver Tank & Mfg. Co. v. *Linde Co.*, 336 U.S. 271 (1949) 11
Green v. *Missouri Pacific Railroad Co.*, 523 F.2d 1290 (8th Cir. 1975) 16
Griggs v. *Duke Power Co.*, 401 U.S. 424 (1971) ..11, 12, 16
Guardians Association of N.Y. City v. *Civil Service Commission*, 431 F. Supp. 526 (S.D.N.Y.), *vacated on other grounds*, 562 F.2d 38 (2d Cir. 1977) 12
Harper v. *Mayor and City Council*, 359 F. Supp. 1187 (D. Md.), *modified and aff'd*, 486 F.2d 1134 (4th Cir. 1973) 12
Pennsylvania v. *O'Neill*, 348 F. Supp. 1084 (E.D. Pa. 1972), *aff'd in pertinent part*, 473 F.2d 1029 (3d Cir. 1973) 12
Pettway v. *American Cast Iron Pipe Co.*, 494 F.2d 211 (5th Cir. 1974) 16
Public Service Comm'n v. *Brashear Freight Lines Co.*, 306 U.S. 204 (1939) 8
Robinson v. *Lorillard Corp.*, 444 F.2d 791 (4th Cir. 1971), *cert. dismissed*, 404 U.S. 1006 (1971) 16
Smith v. *City of East Cleveland*, 363 F. Supp. 1131 (N.D. Ohio 1973), *rev'd on other grounds*, 520 F.2d 492 (6th Cir.), *cert. denied*, 426 U.S. 934 (1976) 12

TABLE OF AUTHORITIES—Continued

Page

United States v. *Bethlehem Steel Corp.*, 446 F.2d 652 (2d Cir. 1971) 16
United States v. *City of Chicago*, 385 F. Supp. 543 (N.D. Ill. 1974), *aff'd on other grounds*, 549 F.2d 415 (7th Cir. 1977) 12
United States v. *Jacksonville Terminal System*, 451 F.2d 418 (5th Cir. 1971) 16
Vulcan Society of N.Y. City Fire Dept., Inc. v. *Civil Service Commission*, 490 F.2d 387 (2d Cir. 1973) 12
Wallace v. *Debron Corp.*, 494 F.2d 674 (8th Cir. 1974) 16
Washington v. *Davis*, 426 U.S. 229 (1976) ..... *passim*

Statutory Provisions:

Civil Rights Act of 1964, as amended by Equal Employment Opportunity Act of 1972, 42 U.S.C. § 2000e *et seq.* ..... *passim*

Guidelines:

Department of Justice Guidelines on Employee Selection Procedures, 28 C.F.R. § 50.14 (1976)

§ 12b(3) 3
§ 12b(5) 6

Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1979) 3

§ 1607.14B(3) 3, 12
§ 1607.14B(6) 6

Miscellaneous:

Standards for Educational and Psychological Tests, 1974, by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education 4

IN THE

Supreme Court of the United States

OCTOBER TERM, 1980

No. 80-406

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA,
Petitioner,

v.

THE UNITED STATES OF AMERICA *et al.*,
Respondents.

On Petition for Writ of Certiorari to the United States Court of Appeals for the Fifth Circuit

---

BRIEF FOR RESPONDENTS JOHN W. MARTIN *ET AL.* IN OPPOSITION

Respondents John W. Martin *et al.* submit this brief in opposition to the petition for a writ of certiorari filed by The Personnel Board of Jefferson County, Alabama.[1]

---

[1] Respondents John W. Martin *et al.* have also, pursuant to Supreme Court Rule 19.5, filed a cross-petition for a writ of certiorari raising the question whether the courts below improperly narrowed the relief granted to respondents.

STATEMENT

The Personnel Board of Jefferson County ("petitioner" or "Board") recruits, screens, and certifies eligible candidates for employment by 12 municipalities and other public agencies, some of whom were defendants below. From December 20 through December 22, 1976, the district court tried the issue of whether petitioner's use of two standardized employment tests violated Title VII of the Civil Rights Act of 1964, 42 U.S.C. §§ 2000e *et seq.* The tests involved are the 10-C examination, used to screen and rank applicants for police officers' positions, and the 20-B examination, used to screen and rank applicants for firefighters' positions. The district court held that petitioner's use of both tests violated Title VII (A. 28-55). It enjoined the Board from making further use of these tests and prescribed goals and timetables to remedy the Board's prior unlawful uses of the tests (A. 64-65). The court of appeals affirmed in respects pertinent here (A. 4-5).

A. Adverse Impact

The district court found that the tests "have had an adverse impact on blacks" as shown by "substantial differences in hire rates and pass rates" for black applicants as opposed to white applicants (A. 33). It found that only 48.6 percent of black applicants passed the 10-C examination, while 90.2 percent of white applicants passed the examination; and that only 24.2 percent of the black applicants passed the 20-B examination, while 85.2 percent of the whites passed the examination (A. 33).

The district court also found that the impact of the test was increased because the Board used the test not only to pass and fail candidates, but also to rank passing candidates for selection on eligibility lists (A. 33, 59). Of the blacks who did pass the tests, 85.4 percent were

placed in the lower half of the police officers' eligibility list, and 89.9 percent were placed in the lower half of the firefighters' eligibility list (A. 33). The district court found that such disproportionately low rankings of black candidates posed an additional barrier to their employment since "in a very real sense, 'passing' an exam is measured not by obtaining a derived score of 70 (and thereby being entered on the eligibility list) but by obtaining a score sufficiently high to be placed on the eligibility list at a position" high enough to result in hiring before the eligibility list expired (A. 59).

In all, the district court found that only 6.6 percent of the black applicants who took the 10-C examination were ultimately hired as police officers, as compared to a 23.3 percent hiring rate for white applicants (A. 32). Of the black applicants who took the 20-B examination, only 3.2 percent were ultimately hired as firefighters, as compared to the 14.1 percent of white applicants thus hired (A. 32).

B. Job-Relatedness

Petitioner attempted to show that the applicants' test scores predicted their job performance through the so-called "criterion" method of test validation, which involves correlating test scores with criteria that must be "important or critical work behaviors or work outcomes." [2]

---

[2] See Department of Justice Guidelines on Employee Selection Procedures, 28 C.F.R. § 50.14 (1976) at § 12b(3), (A. 151), hereafter cited as "Justice Department Guidelines." These guidelines were applicable to the action below since one of the consolidated cases was brought by the Justice Department. Following the trial, the Justice Department, along with other agencies concerned with employment discrimination, joined in promulgating the Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1979), hereafter cited as "Uniform Guidelines" (A. 76-137). The Uniform Guidelines also require that all criteria used in criterion validity studies be "important or critical work behaviors or work out-comes." § 1607.14B(3), A. 101. Accepted

The Board primarily relied on two criteria in this regard: the class ranking of applicants upon graduation from the police and fire academies and their performance on the job as measured by an experimental job rating criterion. (A. 38-40.)[3]

1. *Academy Class Rankings.* The district court concluded that the necessary relationship between test scores and job performance would be established if the tests could predict who would pass and who would fail the academies (A. 38). The court found, however, that the tests could not make such predictions. The court cited evidence showing that

> —the required passing grade for the police officers' examination had been lowered several times, but there was no indication that lowering the score resulted in more candidates failing the training program (A. 46);
>
> —no one ever failed the firefighters' training program (A. 46); and
>
> —if one used available statistical methods to project the academy grades of those scoring below the cutoff points on the examinations, the projection would show that test scores of zero on the examinations would still predict passing averages at the academies (A. 46).

The district court also concluded that correlation of the test scores with academy rankings would be an ac-

---

professional standards also require such a showing. *Standards For Educational and Psychological Tests* (published in 1974 by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education), at 34.

[3] The Board also attempted validation studies using the departments' efficiency ratings of policemen and firemen as a criterion, but the Board's own expert witnesses acknowledged that these ratings were not sufficiently reliable to be an acceptable criteria (A. 39).

ceptable validation if the academy rankings were shown, in turn, "to be themselves appropriate predictors of job performance" (A. 38). While the court found that the evidence showed a statistically significant correlation between test scores and academy grades, it also found that the academy grades did not themselves predict job performance because the Board's studies show

> "that for the most part the correlation between academy grades and measures of job performance are not significant and, in the few instances where significant correlations are found, the findings are mixed—*some being positive and others being negative. A negative correlation, of course, indicates that the higher the academy grades, the lower the performance ratings tend to be.*" (A. 45; emphasis added.)

Thus, the district court found the test scores did not meaningfully predict who would pass and who would fail the academies, and that, although there was some correlation between test scores and academy grades, there was no meaningful correlation between academy grades and job performance. Hence, the district court ruled that the evidence concerning the test scores did not establish that they predicted performance on the job as required by Title VII. The court of appeals explicitly affirmed these findings (A. 21-22).

2. *Experimental Job Performance Rating.* Turning to the evidence concerning the relationship of test scores with the experimental job performance ratings,[4] the district court found that there was no statistically significant correlation between the scores on the firefighter test

---

[4] The experimental job performance ratings were developed for use in the validation studies and consist of 12 rating categories relating to personality characteristics, job knowledge, and abilities. In-service employees were rated by their supervisors on each of these 12 criteria (A. 39-40).

and the job performance ratings (A. 46). Indeed, there was a *statistically significant negative correlation* between test scores and job performance ratings for firemen with more than three years experience, "suggesting that over time the lower scoring applicants made the better employees" (A. 47). As to the police officer test, the district court found that a statistically significant correlation, albeit a rather weak one, existed between the test scores and job performance ratings (A. 47).

The district court then considered whether the correlation between police officer scores and job performance ratings was "appropriate for operational use" by the employer, as required by the Justice Department Guidelines, § 12b(5), (A. 152).[5] The district court considered this question from a number of statistical and practical perspectives (see A. 45-55). It found, *inter alia*, that the observed correlation was so small that a large difference in test scores would predict only a very small increase in job performance ratings and that the correlation could not even tell the Board whether the highest-scoring applicant would likely obtain a "higher experimental rating than the lowest-scoring applicant" (A. 51). In light of "the minimal benefits resulting from the 10-C [police officer examination] in the context of this employment situation," the court concluded that "no-test-at-all has 'substantially' the same practical validity as the 10-C" (A. 54).

---

[5] The Uniform Guidelines, § 1607.14B(6), A. 103, contain the same requirement that an employer show not merely that the relationship between test results and a job criterion is statistically significant, *i.e.*, not likely to be the result of pure chance, but also that it is sufficiently powerful to be "appropriate for operational use" considering, *inter alia*, the magnitude of the statistical relationship, the size of the sample, the value of the criterion used, the degree of adverse impact of the test upon minorities, and the availability of other selection measures.

The court of appeals explicitly affirmed the district court's findings respecting the utility of the test, noting that they were "reached after careful and comprehensive consideration of all the facts and circumstances" (A. 14).

C. *Washington* v. *Davis*

In addition to its review of the evidence supporting the district court's findings, the court of appeals considered petitioner's argument that *Washington* v. *Davis*, 426 U.S. 229 (1976), required reversal. It assumed, without deciding, that *Washington* applied even though that decision rested on the Equal Protection Clause and statutes other than Title VII (A. 16 n. 18). The court of appeals concluded that the test usage sustained in *Washington* was materially different from the test usage condemned by the district court. In *Washington*, the employer had used its test to determine whether a candidate had the minimum reading and verbal skills to complete a job relevant training program. Here, by contrast, the test scores were not used to determine minimal competence to complete a training program and, indeed, did not predict successful completion of the training program (A. 16 & n. 23). Instead, petitioner used the test scores to rank candidates for job selection—*i.e.*, to place candidates on a spectrum ranging from most fit to least fit for the job (A. 16 & n. 23; A. 21).

SUMMARY OF ARGUMENT

Petitioner argues that *Washington* v. *Davis* requires courts to sustain any employment test that predicts class rank in a training program even if the evidence shows that the class rankings have no relationship to job performance. *Washington* v. *Davis* does not support that proposition. There, the test was used simply to screen out applicants who could not read and write well enough to complete the required training program. Here, the

tests were shown to be unable to perform that function. The fact that petitioner's tests nonetheless predicted relative academy rankings does not show that the tests are job-related, because, as both lower courts found from the evidence in this case, there is no relationship between class rank and job performance.

The district court did not, as the petitioner suggests, balance the adverse racial impact of the police officers' examination against its ability to predict job performance, but rather ruled that it had, for practical purposes, no ability to predict job performance. The opinion of the court of appeals further shows that the last question raised by petitioner is not present in the record. In any event, it is proper under Title VII to take into account the adverse racial impact of a selection device when considering if it is appropriate for use by an employer.

ARGUMENT

I. THE DECISION BELOW DOES NOT CONFLICT WITH *WASHINGTON* v. *DAVIS*.

Petitioner's argument that the decision below conflicts with this Court's decision in *Washington* v. *Davis* overlooks important factual differences between the two cases that the court below correctly considered dispositive.[6] It

---

[6] Petitioner presents—but does not argue—the question whether the ruling in *Washington* applies in a Title VII case (Pet. (i)). Petitioner urged the applicability of *Washington* below and effectively prevailed on that issue when the court of appeals agreed to assume its applicability *arguendo*. Petitioner does not now argue that the court of appeals erred in applying the standards of *Washington*. Rather, it merely "suggests" that the Court "may want to address" this "additional issue" should it grant certiorari with respect to any issue both raised *and* argued by petitioner (Pet. 9 n.18). Having prevailed on the issue below, petitioner should not now be heard to raise it as a ground for review of the court of appeals' decision. See *Public Service Comm'n* v. *Brashear Freight Lines, Inc.*, 306 U.S. 204 (1939).

is true that both *Washington* and the present case involve challenges to tests that correlate with performance at training academies. But the differences between the usage made of the tests in the two cases are more important than that superficial similarity, as we now show.

As the court of appeals below pointed out, the test in *Washington* was used to predict (and did predict) which applicants possessed the minimum skills necessary to pass the training academy. See 426 U.S. at 250. It thus screened out persons with insufficient skills to complete training that all conceded was relevant to job performance. *Id.*; see *id.* at 251 n. 17. Moreover, all applicants who passed the test were admitted to the academy (provided that they also met certain character and physical fitness standards). *Davis* v. *Washington*, 348 F. Supp. 15, 16 (D.D.C. 1972). Thus, the decision in *Washington* that the test there at issue was job-related simply represented an application of the common sense principle that an employer may reasonably require that its job applicants read and write well enough to complete a required training program.

Here, by contrast, the tests are not used to predict, and do not predict, which applicants have the minimal skills needed to complete academy training, since statistical projections show that even an applicant who scores zero on the test will pass the academy. Thus, unlike the situation in *Washington*, the tests in this case cannot be characterized—or, justified—as a simple device to screen out those recruits whose reading and writing skills were not sufficient to allow them to complete the required training.

The tests at issue here evidently predict who will make high scores at the training academies, and petitioner argues that they must therefore be upheld as job-related

under *Washington*. In this case, the courts below found, however, that there is no correlation between academy grades and subsequent job performance, and ruled that a test that predicts academy grades that in turn have no relationship to job performance cannot be said to be job-related. Petitioner suggests that the absence of any relationship between grades and job performance is irrelevant under *Washington*, which in petitioner's view held that *any* "positive relationship between performance on [a test] and in [an academy] was sufficient to validate" the test, regardless of the relationship (or lack thereof) between academy performance and job performance. Pet. 11.

That is a misreading of *Washington*, as the court of appeals observed. In *Washington* this Court found it unnecessary, in upholding the test at issue, to determine whether academy performance predicted job performance. But the relationship between academy and job performance was irrelevant for reasons that do not exist here. The test in *Washington* excluded those who would fail the academy. All who passed the test were admitted to the academy, where they received training essential to the job. Relative performance at the academy in those circumstances was unimportant; it was enough to know, to demonstrate that the test was job-related, that the academy training was essential to the job and that those who failed the screening test would not successfully complete the academy course.

In upholding the test, then, this Court did not say that the relationship between relative academy performance and job performance would never be relevant; it held only that it was not necessary to inquire into that relationship when the test in the first instance was shown to exclude those who could not make it through the academy. In the instant case, the tests do not perform

that screening function: one who fails the test would in all likelihood nevertheless pass the academy. Thus the courts below considered evidence regarding the relationship between academy grades—which the tests can predict—and job performance. When it turned out that academy grades bore no relationship to job performance,[7] the courts below correctly held that petitioner had failed to demonstrate that the tests were job-related under Title VII and enjoined petitioner from using the test scores to determine who would be eligible for employment.

That ruling was correct because Title VII forbids the use of selection devices that disadvantage minorities unless the use of those devices is justified by some "business necessity." *Griggs* v. *Duke Power Co.*, 401 U.S. 424, 431 (1971). If a test is shown only to predict relative grades in training, but those grades in training are in turn shown to have no relation to performance on the job, it cannot be said that any "business necessity" permits use of the selection test since, without use of the test, the employer would still achieve the same level of job performance from his work force. This requirement that training grades be themselves shown to predict job performance is simply an application of the general rule, grounded in *Griggs* and contained in all of the guidelines interpreting Title VII, that when an employer seeks to validate employment examinations he must show that the criterion with which he compares test scores itself represents "important or critical work behaviors or work outcomes." See note 2 above. These guidelines "are 'en-

---

[7] Petitioner cites testimony rather than findings below and suggests that there is a relationship between academy grades and job performance. Pet. 8. That suggestion raises a factual issue that was twice resolved against petitioner by the courts below (A. 21-22, 45) and that does not warrant further review by this Court. *Graver Tank & Mfg. Co.* v. *Linde Co.*, 336 U.S. 271, 275 (1949); *Berenyi* v. *Immigration Director*, 385 U.S. 630, 635 (1967).

titled to great deference'" in construing Title VII. *Albemarle Paper Co.* v. *Moody*, 422 U.S. 405, 431 (1975), *quoting Griggs* v. *Duke Power Co.*, *supra*, 401 U.S. at 433-34.[8]

The manner in which the Board uses the examinations provides a further reason why *Washington* v. *Davis* is inapplicable to this case. As the court of appeals noted, unlike the situation in *Washington* v. *Davis*, all applicants who pass the test examination are not automatically admitted to the police or fire academies. Rather,

> "[o]nly those at the top of the eligibility list, those with the highest test scores, are certified for job placement. Even those who score a passing grade,

---

[8] The Uniform Guidelines now specifically require that, when an employer seeks to validate a test using rankings in training as a criterion, relative success in the training program must itself be validated as a measure of job performance. § 1607.14B(3), A. 101-02. Petitioner cites no court of appeals decision which, in conflict with the decision below, rejects this requirement. To the contrary, numerous decisions of the lower courts have required that attempts to validate police officers' and firefighters' examinations against academy rankings must show a substantial relationship between the rankings and job performance. *E.g.*, *Guardians Association of N.Y. City* v. *Civil Service Commission*, 431 F. Supp. 526, 546-47 (S.D.N.Y.), *vacated on other grounds*, 562 F.2d 38 (2d Cir. 1977); *Dozier* v. *Chupka*, 395 F. Supp. 836, 853 (S.D. Ohio 1975); *United States* v. *City of Chicago*, 385 F. Supp. 543, 556 (N.D. Ill. 1974), *aff'd on other grounds*, 549 F.2d 415 (7th Cir. 1977); *Smith* v. *City of East Cleveland*, 363 F. Supp. 1131, 1148-49 (N.D. Ohio 1973), *rev'd on other grounds*, 520 F.2d 492 (6th Cir.), *cert. denied*, 426 U.S. 934 (1976); *Harper* v. *Mayor and City Council*, 359 F. Supp. 1187, 1202-03 (D. Md.), *modified and aff'd*, 486 F.2d 1134 (4th Cir. 1973); *Pennsylvania* v. *O'Neill*, 348 F. Supp. 1084, 1090 (E.D. Pa. 1972), *aff'd in pertinent part*, 473 F.2d 1029 (3d Cir. 1973). *See Blake* v. *City of Los Angeles*, 595 F.2d 1367, 1382 n.17 (9th Cir. 1979), *cert. denied*, 48 U.S.L.W. 3698 (1980); *Vulcan Society of N.Y. City Fire Dept., Inc.* v. *Civil Service Commission*, 490 F.2d 387, 396-97 (2d Cir. 1973).

> and are deemed by the Personnel Board to possess the capacity to complete training, are not hired unless they are among the highest scorers." (A. 21.)

Thus, petitioner's use of the tests for ranking purposes poses an additional barrier to minority employment not found in *Washington* v. *Davis*.

In short, the Board's use of tests is significantly different from the use approved in *Washington*. In that case, the test performed the modest, but critical, task of identifying those candidates who cannot read or write well enough to pass a job training program. The Board's tests, on the other hand, cannot and do not identify those who will fail job training. The Board uses its tests to predict who will be the best students at the academies even though the record shows that there is no relationship between the job performance and the grades of those persons who complete the training programs. *Washington* v. *Davis* does not hold that an employer may use employment tests that have a substantial racial impact solely because the tests predict class standing which has no relationship to performance on the job.

II. THE BOARD'S ARGUMENT THAT THE DISTRICT COURT ERRED IN BALANCING THE RACIAL IMPACT OF THE POLICE OFFICERS' TEST AGAINST ITS VALIDITY DOES NOT WARRANT REVIEW.

A. The District Court Did Not "Balance" The Racial Impact Of The Test Against Its Validity.

The district court found that a weak correlation existed between applicants' scores on the 10-C examination and their ratings on the experimental job performance criterion. Pp. 5-7, *supra*. It then inquired whether this correlation indicated that the examination had any practical utility to the employer. Virtually all of the

court's analysis of this question, which occupied some seven pages (A. 47-54), dealt with techniques used to determine precisely what a correlation of this magnitude would and would not tell the employer. The district court found that, *inter alia*, the test would not even tell the employer that the recruit with the highest examination score would make a better police officer than the recruit with the lowest test score. See p. 6 above. After this extensive analysis, the court in passing mentioned the racial impact of the test.

> "The DOJ Guidelines § 12(b), provide that, in determining operational appropriateness, one should consider the degree of adverse impact of the procedure, the availability of other selection procedures of greater or substantially equal validity and the need of an employer, required by law or regulation to follow merit principles, to have an objective system of selection. Obviously, this latter factor (requirement under civil service law to give some objective test) cannot by itself suffice as justification for a test which has, as here, substantial adverse impact on a racial group. While it can be said that no other available selection [device] with greater validity than the 10-C has been found, yet it must also be said—considering the minimal benefits resulting from the 10-C in the context of this employment situation—that no-test-at-all has 'substantially' the same practical validity as the 10-C." (A. 54; footnote omitted.)

On the basis of this single reference to the Justice Department Guidelines, petitioner argues that the district court "apparently" attempted to "balance the impact of the exam against its validity" and that to do so was error. (Pet. 17). We disagree with that equivocal reading of the district court's opinion. In our view, the court simply adverted once to racial impact in passing and found the test inappropriate for operational use on the

ground that it produced no appreciable increase in the performance of police officers—a finding amply supported in the record. The court of appeals shared that view. In response to petitioner's argument that the trial court balanced validity against racial impact, the court of appeals said:

> "The crucial finding of the district court, which was reached after careful and comprehensive consideration of all the facts and circumstances, is that the validation study showed that the police test is predictive of better job performance, but that the magnitude of the positive prediction is so low that the test is worthless for all practical purposes. We cannot conclude that this finding is clearly erroneous." (A. 14.)

The court of appeals did not say or even imply that the district court balanced racial impact against the magnitude of the correlation. It read the district court's opinion to find that the magnitude of the correlation, standing alone, "is so low that the test is worthless . . ." (*id.*).

In short, the court of appeals did not affirm the alleged balancing of racial impact against the magnitude of a statistical correlation, and so the question presented by petitioner as to the propriety of such balancing does not arise in this case.

B. Title VII Permits A Court To Assess The Racial Impact Of An Employment Practice When Considering Whether It Is Appropriate For Use.

Even if the district court had weighed the racial impact of the 10-C examination when considering whether it was of any practical use to the employer, that would have been permissible. Petitioner cites no decision of this Court or any court of appeals that has accepted its contention that Title VII requires a court to accept any statistical relationship, no matter how slight, between

test scores and a job performance criterion, without taking into account the racial impact of the test.

The Justice Department and Uniform Guidelines specifically authorize consideration of the racial impact of a test.[9] See p. 6 above. In like fashion, numerous decisions of the courts of appeals have taken into account the racial impact of employment practices when considering whether the practices are sufficiently job-related to meet the demands of Title VII.[10] As this Court noted in *Griggs* v. *Duke Power Co.*, *supra*, 401 U.S. at 431, Title VII aims to remove "artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classifications." That purpose would plainly be frustrated if, as the Board urges, an employment practice with the massive discriminatory impact here demonstrated could be justified by any mathematical correlation with any aspect of job performance, no matter

---

[9] Petitioner suggests that the Justice Department Guidelines are not entitled to any deference because the Justice Department is an advocate, rather than an administrative agency with responsibilities in the area of employment discrimination. Pet. 17-18. This ignores the fact that the Justice Department Guidelines were developed and promulgated jointly by that department, the Department of Labor, and the Civil Service Commission. See 41 Fed. Reg. 51734-35 (1976). Moreover, the Uniform Guidelines, which were later adopted by all agencies with administrative jurisdiction over employment discrimination, are identical, in all material respects, to the Justice Department Guidelines.

[10] *E.g.*, *Green* v. *Missouri Pacific Railroad Co.*, 523 F.2d 1290, 1297-98 (8th Cir. 1975); *Wallace* v. *Debron Corp.*, 494 F.2d 674 (8th Cir. 1974); *Pettway* v. *American Cast Iron Pipe Co.*, 494 F.2d 211, 244-45 (5th Cir. 1974); *United States* v. *Jacksonville Terminal System*, 451 F.2d 418, 451 (5th Cir. 1971); *United States* v. *Bethlehem Steel Corp.*, 446 F.2d 652, 662 (2d Cir. 1971); *Robinson* v. *Lorillard Corp.*, 444 F.2d 791, 798-800 (4th Cir. 1971), *cert. dismissed*, 404 U.S. 1006 (1971).

how minimal the correlation or how massive the discrimination might be.[11] Accordingly, even if the district court balanced racial impact and validity and the court of appeals affirmed that determination, further review by this Court is not warranted.

CONCLUSION

The petition for writ of certiorari should be denied.

Respectfully submitted,

RICHARD M. SHARP
WILLIAM R. GALEOTA
SHEA & GARDNER
1800 Massachusetts Avenue, N.W.
Washington, D.C. 20036
(202) 828-2000

Counsel of Record for Respondents John W. Martin et al.

Of Counsel:

WILLIAM L. ROBINSON
RICHARD T. SEYMOUR
National Lawyers' Committee
for Civil Rights Under Law
733 Fifteenth Street, N.W.
Suite 520
Washington, D.C. 20005

Dated: October 15, 1980

---

[11] The Board attempts (Pet. 18-19) to buttress its argument on this point with a hypothetical example having no basis in the record. The calculations performed in this hypothetical depend on the assumption that the increased pool of applicants resulting from an affirmative action program is markedly less qualified than the previous pool of applicants. It is just as realistic to assume that the additional applicants resulting from affirmative action are, on the whole, approximately as well-qualified as previous applicants, but have been previously deterred from applying because of the employer's prior discriminatory practices.

Supreme Court, U. S.
FILED
NOV 21 1980
MICHAEL RODAK, JR., CLERK

Nos. 80-406 and 80-614

In the Supreme Court of the United States

OCTOBER TERM, 1980

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA, PETITIONER

v.

UNITED STATES OF AMERICA, ET AL.

JOHN W. MARTIN, ET AL., PETITIONERS

v.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA

ON PETITIONS FOR A WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE FIFTH CIRCUIT

BRIEF FOR THE UNITED STATES IN OPPOSITION

WADE H. MCCREE, JR.
Solicitor General

DREW S. DAYS, III
Assistant Attorney General

WALTER W. BARNETT
JOAN A. MAGAGNA
Attorneys

Department of Justice
Washington, D.C. 20530
(202) 633-2217

In the Supreme Court of the United States

OCTOBER TERM, 1980

No. 80-406

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA, PETITIONER

v.

UNITED STATES OF AMERICA, ET AL.

No. 80-614

JOHN W. MARTIN, ET AL., PETITIONERS

v.

THE PERSONNEL BOARD OF JEFFERSON COUNTY, ALABAMA

ON PETITIONS FOR A WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE FIFTH CIRCUIT

BRIEF FOR THE UNITED STATES IN OPPOSITION

QUESTIONS PRESENTED

1. Whether Title VII of the Civil Rights Act of 1964 requires a showing of job-relatedness for employment tests which have a substantial disparate racial impact and which are used to select applicants

in rank order for admission into training programs and eventual employment.

2. Whether Title VII prohibits the use of employment tests which have a substantial disparate racial impact, only a small statistical relationship to a job-related criterion, and no operational utility or practical significance.

TABLE OF CONTENTS

	Page
Opinions below	1
Jurisdiction	1
Statement	1
a. Academy performance	4
b. Experimental ratings	5
Argument	9
Conclusion	15

TABLE OF AUTHORITIES

Cases:

Albemarle Paper Co. v. *Moody*, 422 U.S. 405	10, 13
Blake v. *City of Los Angeles*, 595 F.2d 1367, cert. denied, No. 79-54 (Apr. 28, 1980)	10
Comstock v. *Group of Institutional Investors*, 335 U.S. 211	10
Davis v. *Washington*, 348 F. Supp. 15	11
Espinoza v. *Farah Mfg. Co.*, 414 U.S. 86	13
FIRE v. *City of St. Louis*, 616 F.2d 350, petition for cert. pending, No. 80-29	13
Griggs v. *Duke Power Co.*, 401 U.S. 424	10
Guardians Association of the New York City Police Dep't, Inc. v. *Civil Service Commission*, No. 79-7377 (2d Cir. July 25, 1980)	10, 12
Washington v. *Davis*, 426 U.S. 229	3, 7, 10, 11

Constitution, statutes and regulations:

United States Constitution, Fourteenth Amendment (Equal Protection Clause)	3
Civil Rights Act of 1964, Title VII, 42 U.S.C. 2000e	3, 8
Section 713(b), 42 U.S.C. 2000e-12(b)	8, 15

Constitution, statutes and regulations—Continued Page

Department of Justice Guidelines on Employee Selection Procedures, Section 12b(5), 28 C.F.R. 50.14 (1978) 6, 13

EEOC Guidelines on Employment Selection Procedures, 29 C.F.R. (1970):

Section 1607.9 ..... 9, 15

Section 1607.14B(6) 13

OPINIONS BELOW

The opinion of the court of appeals (Pet. App. A-3 to A-27) is reported at 616 F.2d 812. The opinion (Pet. App. A-28 to A-63) and order (Pet. App. A-64 to A-65) of the district court are reported at 13 Empl. Prac. Dec. ¶ 11,504.

JURISDICTION

The judgment of the court of appeals was entered on May 8, 1980. A petition for rehearing was denied on June 16, 1980 (Pet. App. A-2). The petition for a writ of certiorari was filed on September 13, 1980. The cross-petition was filed on October 15, 1980. The jurisdiction of this Court is invoked under 28 U.S.C. 1254(1).

STATEMENT

1. The Personnel Board of Jefferson County has administered employment tests for positions with local government agencies since 1945 (Pet. App. A-28). At issue in this case [1] are two of those tests, Police

[1] This action is a consolidation of four suits brought by the United States and three sets of private parties. (1) The United States' suit was filed on May 27, 1975, against petitioner Personnel Board, Jefferson County, the County Department of Public Health, and 12 cities within the county (Pet. App. A-5 & n.2); (2) on January 4, 1974, the Ensley Branch of the National Association for the Advancement of Colored People and certain individuals filed a class action against the petitioner Board and its Director and the Mayor and City of Birmingham, Alabama (*ibid.*); (3) on January 7, 1974, John W. Martin and other individuals (petitioners in No. 80-614) filed suit against the petitioner Board, Jefferson County, and the City of Birmingham (*ibid.*); (4) on February 20, 1976, Lucy Walker and others filed suit against the

Test 10-C, which has been used since August 18, 1967, and Firefighter Test 20-B, used since October 23, 1968 (*id.* at A-29).

Each test consists of 120 multiple choice questions (Pet. App. A-8).[2] A substantial number of questions on both tests examine primarily for mathematical and verbal ability (*id.* at A-8 n.9). Job applicants who received passing scores were ranked on eligibility lists in order of their scores (*ibid.*). When job openings occurred, the top three candidates on the list were certified to the appropriate hiring department for a final selection (*ibid.*). Once selected, applicants were required to complete a fire or police academy training program before being hired, although, as a practical matter, applicants rarely failed to achieve passing grades in the training program.[3]

Both the firefighter and police tests have had an adverse impact on black applicants. Between March 24, 1972 (the effective date of Title VII for petitioner) and trial, administration of the tests had the following results (*id.* at A-32):

	10-C		20-B	
	Black	White	Black	White
Failing Test	395	191	216	267
Passing Test	373	1,762	69	1,263
Hired	51	455	9	215

---

petitioner Board and the Jefferson County Nursing Home (*id.* at A-5 to A-6) (the police and firefighter tests were not challenged in the Walker suit).

[2] Only 80 of the 120 questions have been used to grade the 10-C test since April 10, 1974 (*id.* at A-8 n.9).

[3] Since 1962, only 11 of 733 cadets have failed the police academy; none has failed the firefighter academy (*id.* at A-45 to A-46).

2. Trial was held in December 1976, limited primarily to the question whether the 10-C and 20-B tests discriminated against black applicants in violation of Title VII of the Civil Rights Act of 1964, 42 U.S.C. 2000e.[4] Petitioner Personnel Board attempted to demonstrate the job-relatedness of the police and firefighter tests by criterion validation studies (Pet. App. A-33 to A-41).[5] The validation studies involved measuring test scores of persons already hired against three criteria: (1) employee efficiency ratings, which are periodic evaluations by superiors of an individual's job performance; (2) grades received in academy training; and (3) experimental ratings developed especially for the validation study to rate job-related personality characteristics, knowledge, and abilities (*id.* at A-38 to A-40). The Personnel Board's own expert witnesses acknowledged, however, that the first of these criteria, efficiency ratings, were "not trustworthy assessments of the employees' actual performance" (*id.* at A-39). Accordingly, the district court rejected these ratings, and considered only the other two.

---

[4] A sex discrimination challenge to the firefighter and police tests and a challenge against an office-worker test were dropped. All other issues were reserved (Pet. App. A-6 n.3). The trial also concerned whether the tests violated the Equal Protection Clause of the Fourteenth Amendment; the district court concluded that use of the tests did not violate the Fourteenth Amendment since "there has been no design or intent on the part of the Board to discriminate on the basis of race or color" (*id.* at A-31).

[5] Criterion validation is one of three recognized procedures for validating employment tests. *Washington* v. *Davis*, 426 U.S. 229, 247 n.13 (1976). "Criterion" or "empirical" validity is "demonstrated by identifying criteria that indicate successful job performance and then correlating test scores and the criteria so identified" (*ibid.*).

a. Academy Performance

With respect to academy grades as a criterion, the district court stated that "[w]here, as here, new employees are required to complete special training before performing their duties, successful completion of that training may properly be used as a criterion-measure, if, that is, the training is intended to, and does, provide skills or knowledge needed for performance of the job" (Pet. App. A-38). The court concluded that the two schools here "do serve that purpose and function" (*ibid.*). But the district court found that neither the 10-C nor the 20-B test validly predicted successful *completion* of the required training courses (*id.* at A-45). Indeed, the court found that even persons scoring zero on the 10-C and 20-B tests could be predicted to achieve passing academy grades (*id.* at A-46).

Although test scores were shown to have a statistically significant relationship to academy grades, the district court held that relative standing or ranking among students who successfully complete the training program was not a proper validating criterion absent a showing that relative rankings or standings are, in turn, "appropriate predictors of job performance" (Pet. App. A-38). The court expressed two concerns that cautioned against accepting academy ranking as a relevant criterion: (1) because academy grades have no impact on job opportunities after completion of the training program, they are "likely to be influenced by motivational considerations not present in actual job performance"; and (2) the "emphasis in the academies on paper-and-pencil multiple choice items, while providing objectivity, may also reflect a relationship to the paper-and-pencil

screening exam not found in job performance" (*id.* at A-39). The district court found that while test scores correlated with academy grades, the correlations between academy grades and subsequent job performance were, for the most part, not significant (*id.* at A-45). And where significant correlations *were* found, the findings were mixed—some were positive, but others were, in fact, negative (*id.* at A-45). A negative correlation, of course, indicates that the higher the academy grades, the *lower* the performance ratings.

b. Experimental ratings

Experimental ratings were deemed to be an appropriate criterion measure, although the district court found them limited in their usefulness because of "lack of evidence as to reliability and the lack of special steps to detect or control possible bias" (Pet. App. A-40). In the present case, the studies relating to experimental ratings and the 20-B test were found to have produced inconclusive results. Thus, while there was a positive correlation between test scores and job performance of firefighters on the job less than three years, there was a negative correlation between 20-B test scores and job performance of firefighters with more than three years experience—"suggesting that over time the lower scoring applicants made the better employees" (*id.* at A-46 to A-47). Thus, the court concluded, "[a]bsent any indication that during the first few years the lower scoring applicants had been inadequate on the job, one is hard pressed to conclude that the higher scoring 20-B applicants are in fact the better persons to hire" (*id.* at A-47).

With respect to the 10-C test, the court, after undertaking a number of statistical analyses, reached

the conclusion that although the test had a statistically significant positive relationship to the experimental ratings, "albeit of very low magnitude" (Pet. App. A-53), the test had only marginal utility as a selection device for improving police officer performances (*id.* at A-47 to A-53). The court also included in its analysis several practical considerations: (1) the nature of the job and the consequences of a faulty hiring decision (*i.e.*, that the occasional selection of an inept police officer would not present the same type of daily economic and human risk factor as would, for example, the employment of inept airline pilots or busdrivers) (*id.* at A-53); (2) the fact that new police employees must undergo a 12-month probationary period during which incompetents could be weeded out (*id.* at A-54); (3) the fact that this test did not address the more critical need of police forces to screen out the emotionally unfit (*ibid.*); and (4) the substantial adverse racial impact of the test (*id.* at A-54 to A-55).[6] Finally, the district court concluded that the benefits of the 10-C test were so "minimal" that "no-test-at-all has 'substantially' the same practical validity as the 10-C" (*id.* at A-54).

Based on its conclusion that the 10-C and 20-B tests had an adverse impact on black applicants and had not been shown to be job-related, the court found that they violated Title VII. The district court concluded that petitioner's liability for the use of the 10-C and 20-B tests commenced on April 25, 1975, and July 8, 1976, respectively, the dates on which

---

[6] The court relied on Section 12b(5) of the Department of Justice Guidelines on Employee Selection Procedures, 28 C.F.R. 50.14 (1978) (Pet. App. A-52), which provides that the degree of adverse impact should be considered in determining whether a selection procedure is appropriate.

petitioner received the validation study results and "should have * * * concluded that provisional use of the tests was no longer permissible" (Pet. App. A-56).

3. Petitioner Personnel Board appealed from the ruling that the tests violated Title VII. The United States and the Martin plaintiffs cross-appealed on the question of the dates for commencement of liability.[7] The court of appeals affirmed the district court's holding that the 10-C and 20-B tests violated Title VII but remanded for additional findings on the question of when liability commenced.

Applying the "clearly erroneous" standard, the court of appeals affirmed the district court's findings that the tests were not shown to be properly validated as predictive of job performance (Pet. App. A-13). The court of appeals rejected petitioner's argument that the tests were properly validated under Title VII because test scores correlated positively with training academy grades (*id.* at A-14). The court distinguished *Washington* v. *Davis*, 426 U.S. 229 (1976), where this Court held that an employment test was properly validated because it was shown to predict that those tested had the minimum reading and verbal skills necessary to complete a job-relevant training program (Pet. App. A-15 to A-16). Here, by contrast, the court found that the tests were not used to test *minimum* skills, "but rather were used to rank job applicants according to their test scores and to select only the highest test scorers for job placement" (*id.* at A-16).[8] Indeed, the court of appeals

---

[7] The district court's determination that there was no constitutional violation was not challenged on appeal.

[8] Because *Davis* was distinguished in the manner indicated in the text, the court of appeals declined to decide whether

concurred (*id.* at A-21 & n.23) in the district court's finding that the tests were "inappropriate even for * * * determining whether an applicant has the minimum capacity to complete training" (*ibid.*) because a score of zero on the tests appeared to "predict passing academy averages" (*id.* at A-46).

The court of appeals declined petitioners' invitation to extend the *Davis* rationale to hold that *any* test can be validated on the basis of its relationship to training "without regard to the test's ability to predict job performance" (Pet. App. A-14), because "[s]uch an extension would violate the requirement of job performance validation enunciated in *Griggs* and *Albemarle*, as well as the agency guidelines elaborating upon that requirement" (*id.* at A-20; footnote omitted). The court held that ranking on the basis of entry level test scores was permissible only upon a showing that higher test scorers performed better on the job than lower test scorers (*id.* at A-21 to A-22), and that such evidence was "utterly lacking" here (*id.* at A-22). The court of appeals accordingly affirmed the district court's judgment that petitioner's use of the 10-B and 20-C tests violated Title VII.

The case was remanded for additional findings relating to the dates on which liability commenced. The court of appeals directed that petitioner Board should be held immune from liability, pursuant to Section 713(b) of Title VII,[9] during the period when

---

the *Davis* standards applied to this, a Title VII case (*id.* at A-16 n.18).

[9] Section 713(b) of Title VII, 42 U.S.C. 2000e-12(b) provides, in pertinent part:

> In any action or proceeding based on any alleged unlawful employment practice, no person shall be subject to any liability or punishment for or on account of (1) the commission by such person of an unlawful em-

its validation studies were being conducted only if it in good faith relied on and met the conditions for provisional use of tests provided in Section 1607.9 of the EEOC Guidelines (Pet. App. A-25 to A-27).[10]

ARGUMENT

The district court and court of appeals properly concluded that the tests at issue had not been shown to be job-related for Title VII purposes. The fact that high test scores correlated with high grades in the academy training programs was properly deemed insufficient for validation purposes absent a showing that the high test scores or high academy grades predicted better job performance. The soundness of this holding is underscored by the fact that in this case, "even extremely low test scores do not predict failure at the academies" (Pet. App. A-12). In light

---

ployment practice if he pleads and proves that the act or omission complained of was in good faith, in conformity with, and in reliance on any written interpretation or opinion of the [EEOC] * * *.

[10] Section 1607.9 of the EEOC Guidelines on Employment Selection Procedures, (29 C.F.R. (1970)) (Pet. App. A-180) provides:

> Under certain conditions, a person may be permitted to continue the use of a test which is not at the moment fully supported by the required evidence of validity. If, for example, determination of criterion-related validity in a specific setting is practicable and required but not yet obtained, the use of the test may continue: *Provided*, (a) The person can cite substantial evidence of validity as described in § 1607.7(a) and (b); and (b) he has in progress validation procedures which are designed to produce, within a reasonable time, the additional data required. It is expected also that the person may have to alter or suspend test cutoff scores so that score ranges broad enough to permit the identification of criterion-related validity will be obtained.

of the concurrent findings of the two courts below on the invalidity of the tests (see *Comstock* v. *Group of Institutional Investors*, 335 U.S. 211, 214 (1948)), this case presents only a particular application of well-settled legal principles. The decision below is consistent with the decisions of this Court and those of other courts of appeals.

1. Petitioner does not dispute that the tests have a substantial disparate impact on black applicants. Where there is such an impact, Title VII prohibits the use of such a device absent a "manifest relationship to the employment in question." *Griggs* v. *Duke Power Co.*, 401 U.S. 424, 432 (1971); see also *Albemarle Paper Co.* v. *Moody*, 422 U.S. 405 (1975). Petitioner nevertheless asserts that this Court's decision in *Washington* v. *Davis*, 426 U.S. 229 (1976), permits an employer to validate a test for Title VII purposes by establishing a positive relationship between the test scores and a training program without regard to whether the test scores themselves have any relationship to job performance.

Washington v. *Davis*, of course, was not a Title VII case, and it is by no means clear that it applies in a Title VII context.[11] However, as the court of appeals properly concluded (Pet. App. A-16 n.18), that question is not presented because the holding in *Davis* is inapplicable here in any event.

The principal distinction is that the tests here were used for *ranking* as well as for *screening* of job ap-

---

[11] See, *e.g.*, *Washington* v. *Davis*, *supra*, 426 U.S. at 255 (Stevens, J., concurring); *Blake* v. *City of Los Angeles*, 595 F.2d 1367, 1378-1382 & n.17 (9th Cir. 1979), cert. denied, No. 79-54 (Apr. 28, 1980); *Guardians Association of the New York City Police Dep't, Inc.* v. *Civil Service Commission*, No. 79-7377 (2d Cir. July 25, 1980), slip op. 6087-6088.

plicants. In *Davis*, police recruits were in great demand, and all candidates who passed the exam were offered jobs. *Davis* v. *Washington*, 348 F. Supp. 15, 16 (D.D.C. 1972). Here, in marked contrast, only a small percentage of the hundreds of persons taking the tests were ultimately offered jobs. Thus, only 6.6% of black applicants and 33.3% of white applicants who took the 10-C test were hired, and only 3.2% of black and 14.1% of white applicants taking the 20-B test were hired (*ibid.*).

A further difference is that in *Davis* there was substantial evidence that those who failed the test would have been unable to complete the Recruit School successfully and, as a result, that there was a " 'direct and rational relationship between the content and difficulty of Test 21 and successful completion of recruit school training.' " 426 U.S. at 251 n.17. In other words, there was evidence in *Davis* that "the level of verbal ability measured by Test 21 is, at most, no higher than that required for successful completion of recruit school training as determined by the final grade average attained by the recruit" (U.S. Supp. App. 13)[12] and that persons with the lowest scores on Test 21 would probably have obtained failing grades in the recruit training program (*id.* at 16-17). By contrast, in this case, there is evidence that even those who failed the entry-level tests would have been able to complete academy training successfully (Pet. App. A-12, A-46). In any event, many applicants who received passing scores on the 10-C and 20-B tests, and thus, by petitioner's own standard, possessed the necessary qualifications to complete

[12] "U.S. Supp. App." refers to the United States Supplemental Appendix filed in the court of appeals in this case, a copy of which is being lodged with the Clerk.

training, were not offered jobs because eligibility lists for hiring were ranked in order of test scores.

For these reasons, the court of appeals properly declined to extend the *Davis* holding to this case "where the tests were not used to ascertain the minimum skills necessary to complete job-relevant training, but rather were used to rank job applicants according to their test scores and to select only the highest test scorers for job placement" (Pet. App. A-16). *Washington* v. *Davis* does not provide employers with a vehicle to dispense with the requirement mandated by this Court in *Griggs*, *supra*, and *Albemarle Paper Co.*, *supra*, that employee selection devices that impact adversely on blacks must be shown to have "a manifest relationship" to the jobs in question. "Such evidence is utterly lacking here" (*id.* at A-22).

The Fifth Circuit's holding in this case is consistent with those of other courts of appeals. *Guardians Association of the New York City Police Dep't, Inc.* v. *Civil Service Commission*, No. 79-7377 (2d Cir. July 25, 1980), presented a situation similar to this case. The plaintiffs were black and Hispanic police officers who had been laid off under a last-hired, first-fired policy. They challenged the validity under Title VII of Entry-level employment tests used, as in this case, to rank candidates according to test scores. As here, candidates were then hired in rank order. The use of the test as a screening device was not in issue in *Guardians*, because all of the plaintiffs had passed the test. "They challenge[d] only the use of their passing scores to determine the order in which they were offered employment" (slip op. 6089-6090). The Second Circuit concluded, as the Fifth Circuit did here, that even if *Washington* v. *Davis* applied to Title VII cases, "the *Washington* v. *Davis* argument—that the

entry-level tests were justified by the need to weed out those applicants who lacked the minimal skills necessary to successful completion of the training program —is simply not available to the defendants on this appeal" (*ibid.*). Similarly, in *FIRE* v. *City of St. Louis*, 616 F.2d 350 (8th Cir. 1980), petition for cert. pending, No. 80-29, the Eighth Circuit held that a Fire Department promotion test used as a ranking device could be validated for Title VII purposes only by showing that the higher ranking test scorers made the better employees. 616 F.2d at 357-360.[13]

2. The lower courts properly required a showing of practical or operational significance for the use of the 10-C test, and the district court properly considered, among other factors, the degree of adverse impact of the 10-C test. Petitioner does not challenge, as such, the requirement that a showing of practical significance must be made under Title VII, but urges that it is not appropriate to balance a test's practical utility against its adverse impact (Pet. 16-20). Petitioner therefore suggests that this Court sanction a departure from the requirement of such a balancing determination, which is set forth in the Department of Justice Guidelines on Employee Selection Procedures, Section 12b(5), and the EEOC Guidelines on Employment Selection Procedures, Section 1607.14 B(6). However, this Court has ruled that the Guidelines, though not binding, are entitled to great deference. *Albemarle Paper Co., supra,* 422 U.S. at 431. The only exception or limitation to such deference is where "there are 'compelling indications' " (*Espinoza* v. *Farah Mfg. Co.*, 414 U.S. 86, 94-95 (1973)) that

---

[13] The test's relationship to a training program was not at issue in the St. Louis case.

the administrative interpretation is inconsistent with congressional intent. Petitioner has neither demonstrated any such compelling indications here nor otherwise suggested why the criteria used by the courts below are not both reasonable and appropriate in this context. Rather, petitioner's only objections are theoretical ones.

Petitioner hypothesizes that the Guidelines' requirement of considering a test's adverse impact in assessing its practical utility will have a chilling effect on affirmative action programs (Pet. 18-19). The suggestion is that such programs would attract a large number of minority applicants who are disproportionately unqualified when compared with those attracted to the jobs in the ordinary course of business and, thus, that an employment test used in such a circumstance might have a substantial adverse impact on minorities even though it is job-related. Whether, in some circumstances, this theory might have merit, it is not a valid basis for challenging the lower courts' decisions here, "[a]bsent * * * any hard data to support such an hypothesis or to indicate its magnitude" (Pet. App. A-33).[14]

---

[14] Similarly, petitioner's additional contention that balancing adverse impact against practical utility may restrict the use of nationwide tests because adverse impact may differ from region to region (Pet. 19-20) is without a factual basis in this record.

CONCLUSION

The petitions for a writ of certiorari should be denied.[15]

Respectfully submitted.

WADE H. MCCREE, JR.
Solicitor General

DREW S. DAYS, III
Assistant Attorney General

WALTER W. BARNETT
JOAN A. MAGAGNA
Attorneys

NOVEMBER 1980

---

[15] Cross-petitioners argue that the court of appeals incorrectly interpreted Section 713(b) of Title VII to allow immunity for petitioner if it can establish that it in good faith relied on, and complied with, EEOC's guideline relating to the provisional use of tests.

This question does not merit the Court's review at this time for two reasons. First, as cross-petitioners note (Cross Pet. 10-11 n.7), the Employee Selection Procedure guidelines promulgated since 1976 make quite clear that employers are not immunized from liability for tests in use during the period in which validation studies are being conducted. The question thus arises only with respect to tests used prior to November 1976.

Second, the court of appeals' ruling may be of no practical significance even in this case. The Personnel Board may be unable to meet the court of appeals' requirement (Pet. App. A-25 to A-27) that it in good faith relied on Section 1607.9 or that it complied with its provisions—*i.e.*, that it had substantial evidence of the tests' validity and that its validation studies were designed to produce, within a reasonable time, the additional validation data required. If the Board cannot make such a showing, cross-petitioners are not injured, of course; if the Board does make the required showing, cross-petitioners will have an opportunity to seek this Court's review of the question at that time.

☆ U. S. GOVERNMENT PRINTING OFFICE: 1980 333198 150

see

80

614